


A Call for Keeping Doors Open and for Parallel Efforts

Paul De Boeck^a , Jolynn Pek^a, Katherine Walton^a, Duane T. Wegener^a, Brandon M. Turner^a, Barbara L. Andersen^a, Theodore P. Beauchaine^b, Luc Lecavalier^a, Jay I. Myung^a, and Richard E. Petty^a

^aDepartment of Psychology, The Ohio State University Columbus, Ohio; ^bUniversity of Notre Dame, Notre Dame, Indiana

We are grateful for the commentary papers and appreciate the comments, the critical discussions, and disagreements. In our responses, we clarify important points and further the ongoing discussion, hoping that this discussion about constructs that began in this journal can continue on other fora and in follow-up exchanges with interested scholars. We discuss key points made in the commentaries and argue for keeping doors open for different approaches while pursuing corresponding efforts in parallel, rather than to focus on a single explanation of the problems psychology as a science is confronted with.

To briefly summarize, De Boeck et al. (this issue), which we term our Questioning paper, introduced a controversy in relation to explanations for some fundamental issues in psychology. These issues include observed heterogeneity and inconsistencies in psychological results, which are linked to dissatisfactions with broad and abstract constructs that reflect the complexities of psychological phenomena. In response, some methodologists suggest improving operational definitions (ODs) of constructs and improving measurements (called upscaling ODs in the Questioning paper) and also using more specific constructs (called downscaling constructs in the Questioning paper).



Instead of attributing observed variability in psychological results to suboptimal conceptualizations (e.g., type of constructs) and methodological practices (e.g., less careful measurement), we consider an alternative approach. We revisited the nature of constructs to better align constructs with a variable and complex psychological reality. In particular, we posit that constructs should be modified to be more flexible than what is commonly considered as desirable for constructs. Stated differently, even if we were to measure all persons within a target population (reducing sampling variability to 0), tighten measurement to an idealized level, and ensure that all statistical practice is honest—reducing questionable research practices (QRPs), researcher degrees of freedom (RDF), p-hacking—to zero, we conjecture that there would still remain variability and uncertainty inherent to a complex and variable reality.

Therefore, we propose considering more flexible constructs and accompanying methodologies. These methodologies are based on accepting measurement continuity and

generalization principles. Measurement continuity means that, apart from sampling variation, there always will be variation in the measurement of constructs. Generalization approaches can formally model the variability so that inferences can be made under a more realistic degree of uncertainty. Thus, contrary to promoting a perceived loosening of methodological rigor, we advocate the use and development of methods by an acknowledgement of the uncertain nature of constructs.

Although we have created some controversy in our Questioning paper, it was – as indicated – for strategic reasons. We do not believe there is an either/or choice to be made in approaching psychological research. Our main motivation was to consider an alternative for the growing consensus that abstract constructs and measurement practices are the problem of today's psychology. How can we be so sure that the growing consensus (in line with upscaling ODs and downscaling constructs) is the best option? An alternative view is that psychological phenomena are too complex and variable to be sufficiently captured with the available or idealized methods and conceptualizations. Betting on just one or a few options might not be wise. Thus, we prefer to keep doors open and invest in parallel efforts. If there is a one best approach, it will show, and if there is not, that is informative as well.

Our responses begin with preliminary points to clarify (in response to Uher) how we believe the typical psychologist considers constructs and uses terms related to measurement. This sets the stage to respond to other comments received. In general, points were raised about: (a) the role of constructs and theory (Haig; González and Irribarra; Machery); (b) methods as a source of variability in empirical findings (Bollen; Hodson; Machery); (c) operational definitions and the multi-referential nature of psychological constructs (Uher); and (d) the role of epistemological theories to interpret the issues we focus on in our Questioning paper (Haig; Uher). We are aware that the topics we discuss are only a selection of issues raised, and there are many other valuable important ideas and arguments in the commentary papers that are worthy of further discussion.

CONTACT Paul De Boeck  deboeck.2@osu.edu  Department of Psychology, Lazenby Hall, Ohio State University, 1827 Neil Avenue, Columbus, OH 43210.

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Preliminaries

To clarify the terminology used in our Questioning paper and to avoid new misunderstandings in our responses, we explicate our assumption that psychologists are realists in an implicit way when they build theories and use constructs as building blocks. We also prefer to assume, in response to Uher, that most psychological researchers think of constructs as (i) existing or hypothetical *properties* (treated in research as if they exist) with (ii) a definition, which explicates the *concept* of a property, and (iii) with a label, a *term* to denote the property and the concept. Uher uses the term referent for the property, construct for the concept of the property, and signifier for the term. Her criticism of psychology and our Questioning paper is that often, one or more of the three pairs of the triad are conflated. This triad of (i) to (iii) nicely agrees with standard terminology in metrology published by the International Organization for Standardization (ISO 4th edition 2022). The triad is also clearly explained by Mari et al. (2019) who used a different formal notation for each of the three and formulated their theory of measurement across the sciences relying on this triad.

Properties of objects are being measured in psychology. The objects can be persons or events (e.g., processes) and the values of the properties can vary between and within objects. These properties are the attributes in the Cronbach and Meehl (1955) definition of constructs, although Cronbach and Meehl focus on persons as objects in their psychometric approach. In the psychological literature these properties are also called variables and in a statistical model, they are called latent variables, because they are not directly observable and can be approached only through manifest (i.e., observable) variables, also called indicators. Psychologists and many other scientists use the term “construct” to designate both, the property and the concept of the property. It is also a common practice that psychologists write in a way seemingly to identify a term (a label) with a (hypothesized) property that is measurable in some way and has explanatory value. It is clear that a term alone cannot have an explanatory value; instead, it is the property to which the term refers that can have an explanatory role.

For a number of reasons, we feel reassured in the opinion that psychologists differentiate between concept, property and term, rather than conflate those three, at least in an implicit way. When researchers use a *term* in their papers, they tend to give a definition (i.e., the *concept* indicated by the term), and they use operational definitions of constructs as the observable indicators of the *property*. There might be issues with definitions such as definitions being non-committal or the unavailability of rules to derive operational definitions from conceptual definitions other than vague requests for the alignment of conceptual definitions and operational definitions. Additional potential issues with definitions include the implicit and possibly false belief that the same term necessarily implies the same concept and property (jingle fallacy) and that different terms necessarily imply different concepts and properties (jangle fallacy). One would not feel the need to define a term and to select an operational definition if one would not differentiate between terms, concepts, and properties as referents. Empirical

research is set up to investigate the hypothesized explanatory role of the (hypothesized) properties.

Perhaps the most critical point Uher makes is that psychologists, and we in our Questioning paper, commit the sin of conflation. The language that many psychologists use, including ourselves, might be confounding at times, though we still understand each other rather well. Examples of confusing language do not prove an absence of conceptual differentiation. So many linguistic expressions do not make logical sense when interpreted in a literal way, whereas they are clear for readers and writers. The middle voice in English is a simple example. For example, “The bass notes do not hear very clearly” (Halliday, 1967, p. 49). It does not make sense that notes can hear. See Davidsen and Olivier (2008) for many more examples. Let this not be an excuse, it is only an explanation of how text can be misleading depending on how it is interpreted. In response to Uher’s invitation to self-criticism, we admit that confusion (including any we have created) can lead to logical errors. In what follows, we hope that our language is clear from the context. The remaining topics we have selected from the commentaries will be discussed next.

Theory and Constructs

In our Questioning paper we considered a theory as a set of hierarchical and/or explanatory relationships between the constructs and between components within a construct (i.e., between the properties). The distinction between theory and construct is not always very clear in that the meaning of a construct depends on its relations, in that a construct may have components with theoretical relations, and in that constructs have an explanatory role for other constructs. The commentaries raised two questions about the role of constructs and theory. Haig’s commentary questioned our focus on constructs in the context of theories, and commentaries by González and Iribarra and by Machery concerned the location of the variability and complexity of psychological findings, which can be in the constructs, or in the theory (relation between constructs), or both.

Haig in his commentary proposed to forego the role of constructs and to replace the gap between constructs and their ODs (the C-OD gap) by the gap between observations and theory. His recommendation is to fill the gap between theory and observations with psychological phenomena. Filling the gap in this way would make our focus on constructs a distraction. The way we understand Haig’s proposal is that he prefers a bottom-up approach with theories built from psychological phenomena. In his 2008 *Psychological Methods* paper, Haig (2005) described the approach as guided by abductive principles, not by constructs. So why care about constructs? “Mechanistic” theories (i.e., causal process theories) as explained in his commentary and in Haig (2005, p. 272) are little concerned with definitions of constructs. Such an outlook is “at variance” (Haig’s term) with the focus on constructs in our Questioning paper. We believe that Haig’s view is interesting and realistic as a view on the context of discovery though most articles in

psychology are written from a context of a justification perspective. Alternatively, one can consider Haig's suggested approach as exploratory (discovery) versus the still common ideal of a (confirmatory) hypothesis testing approach (justification).

Our interpretation of Haig's proposal may be too simplistic in light of a further elaboration by Haslbeck et al. (2022). In Haslbeck et al. (2022) abductive approach, they started with target phenomena, for example, three panic disorder phenomena (recurrent panic attacks, persistent concern, and avoidance behavior) and a formal model of a generic kind (such as a network model), to simulate data to assess how they fit with empirical data. Based on possible discrepancies between simulated and empirical data, the model is adjusted, and a panic disorder theory is developed. Because there are prescriptive and testing (confirmatory) steps involved in the approach, one can hardly call it a purely explorative approach. To us, the three panic disorder phenomena imply low-abstraction constructs (i.e., rather concrete constructs; see the Zeigarnik effect example in our Questioning paper). We doubt that one can do without constructs.

Haig in his commentary paper also refers to the Flynn effect as a psychological phenomenon in which IQ increases across generations. Knowing how IQ is measured in different ways, we consider IQ as an abstraction, and because IQ is used in other publications as an explanatory variable, it qualifies as a construct. Certainly, the notion of intelligence in the Dickens and Flynn (2001) theory of intelligence, to which Haig refers, is a construct. Perhaps the role of these constructs is secondary in the abductive process and less important than the phenomenon of increasing IQ. In sum, while it does make sense to establish psychological phenomena and to build theories without necessarily starting from constructs, this approach still needs constructs to make sense of phenomena and to develop theories. In the Questioning paper we used a top-down approach, so we are happy that Haig points to an alternative more bottom-up approach, with possible hypothesis testing components as in Haslbeck et al. (2022). The role of constructs might be reduced by taking an abductive approach, but in such an approach the complexity and variability issues have moved to become variation in the foundational psychological phenomena.

Other comments on theory and constructs concern where the varying character of psychological phenomena is situated, either in the theory (relations between properties) or in the properties themselves. If the relation between X and Y as properties varies, either X and Y are fixed properties with variable relations, or X and Y are varying properties (which would manifest as variable relations in empirical studies). In our interpretation, González and Iribarra and Machery believe the former – if the empirical relations between X and Y vary, the variation can (and should?) be explained by the complexity of theories; for example by extending the theory with moderators. It may be difficult to differentiate between the two views: varying constructs or varying relations, because these two views map onto the same heterogeneity of findings.

However, the two views have different implications. According to the varying constructs view, the differences would often result in lack of measurement invariance. According to the varying relationships view, however, it should often be possible to reach measurement invariance, perhaps after refining definitions and implementing improved measurement methods, and then identifying consistent moderators of the X-Y relations. The moderation approach is quite standard but may not be feasible if the moderations would be highly complex.

In his commentary, Bollen raises the interesting question of whether our view is falsifiable. We believe it is; if not now, then certainly in the long run. If the case of temperature turns out to apply to psychological constructs, as suggested by Machery that it might, then our view is falsifiable on the condition that sufficient efforts are made to improve definitions and measurement in the line as advocated, for example, by Flake and Fried (2020). We are not against such steps; instead, we prefer to keep doors open, while working in parallel on methods to deal with varying findings, whether due to the complex psychological reality or to various shortcomings of current research practices.

Methods as a Source of Variability

Two different broad categories of sources of variability in research findings will be discussed here. We do not refer to sampling variability in the common sense of sampling from a population of persons, nor do we refer to imperfect reliability or measurement error. They are evident and well-known sources of variation with established ways to deal with those sources. We will instead contrast (a) variability due to methodological practices and (b) variability as a reflection of the psychological reality. Variation stemming from methods refers to differences between methods that are used as well as suboptimal methods, including insufficiently refined definitions and inadequacies in designs and measurements. Method-based variation, including QRPs/RDFs/ p-hacking, is what Bollen, Hodson, and Machery have in mind to explain heterogeneity in psychological findings. Similar to the history of defining and measuring temperature, Bollen, Hodson, and Machery argue that we need to keep working hard and have patience so that the problem will eventually be solved. This is a reasonable and laudable view, which we also encourage readers to pursue as a possible solution. Many of us, the coauthors of the Questioning paper, do largely invest in such efforts in their own research. The weakness in banking on this single perspective is assuming that such efforts would guarantee success. But what if these efforts do not completely resolve the heterogeneity issue? We prefer to keep the door open to a possibility that implies parallel efforts to deal with the variation and the associated uncertainty. To be clear, variation is a finding, whereas the associated uncertainty matters for the inferences made from such findings.

The source of variation we contrast with methods is the psychological reality along with the constructs used to understand, explain, and predict that reality. We used the

term “measurement continuity” to refer to the resulting continuous variation of measurement outcomes. One can argue that there should be explanations (e.g., moderators) for variabilities such as violations of measurement invariance. Machery is explicit about his belief that explanations need to be sought. Explanations would contribute to psychology as a science, but we also believe there is not always an explanation that is feasible or even desirable; such explanations might require unraveling huge complexities. Such an approach to simplifying complexity is in fact already accepted for other forms of complexity. For example, it is common practice to collapse higher-order interactions in analysis of variance and to add them to the residual variance. Other examples include error terms for imperfect reliability (e.g., of tests) and for incomplete explanation (e.g., in a regression model). We argue that the concept of adventitious error (Wu & Browne, 2015) fulfills a similar simplifying role. Adventitious error is varying “bias” if one believes that the property a construct refers to is fixed. The random error notion is used in the literature to deal with large sets of disturbances, although in theory, one might be able to interpret the disturbances in a meaningful way (e.g., due to moderators). In fact, error terms are a form of entification of a statistical construction, as if the error really exists and has causal influence on observations. For reasons of parsimony (to simplify the enormous complexity), we typically follow Spearman’s genial idea to treat the disturbances and specific terms as sources of variation independent of the property we are interested in, *so that* one can isolate these nuisance disturbances from the focal property. If desired, we can correct for the attenuations (of parameter estimates) induced by imperfect reliability, which nowadays means using latent variables or forms of hierarchical Bayesian modeling. The “so that” in the present paragraph is a guaranteed “so that” because it follows from the definition of random error terms. The resistance against varying constructs in the sense of properties is surprising in the light of the easiness of accepting the notion of unexplained error.

Why should we not think in a similar way about validity as we do about reliability? If the nature of constructs is complex and also varies, one may expect two consequences. There would not be perfect convergence between two sets: (a) between different measures of the same construct, and (b), between the relations of measures used in different studies for the same construct (which would show up as violations of measurement invariance). That is, for complex constructs such as depression one may expect different selections of content elements used as indicators for the construct, without high convergence between these different selections. Also, if the nature of a construct varies, one would expect varying relations within the same set of indicators, which would show up as a violation of measurement invariance. In both cases the latent variable as a statistical representation of the property will move within a not strictly delineated area, just like random error varies and is treated as unbounded. For the same set of indicators, this kind of randomness of relations beyond sampling variation, is called adventitious error by Wu and Browne (2015) and discussed

as a fundamental notion in our Questioning paper. If one considers a property as fixed as in the bull’s eye metaphor, then the variation around the bull’s eye would be bias (i.e., deviation from the bull’s eye) and the source of imperfect validity, just as the error term is the source of imperfect reliability.

Measurement continuity around a common core can be interpreted as random bias across studies. In our view, measurement continuity might not be bias but it reflects the varying nature of constructs (i.e., of the property). Adventitious error offers an elegant approach for a kind of variation that would ideally be explained through a large set of moderators and complex interaction effects (just as error terms – i.e., variation in a relation across participants could be explained in theory). Adventitious error makes the latent variable land into somewhat different positions in the geometric representation of the latent variable model. Together these different positions of the latent variable define an area (instead of a point; see Figure 1 in our Questioning paper), which results in increased standard errors of parameter estimates. A very nice feature of this statistical conceptualization is that the RMSEA goodness of fit measure can be used for an estimate of the increased standard error quantifying the width of the measurement continuity area that statistically represents the construct (see also Wu & Browne, 2015). Adventitious error, as one type of conceptualization of construct variation, is one of the reasons to represent constructs as an area in which the estimated latent variable can land, not because of sampling variation but, instead, due to adventitious error. Constructs as areas is not a vague metaphor, unless one considers statistical models that explicitly model different sources of variability as vague and purely metaphorical.

A consequence of the view that a construct represents an area is that related constructs are no longer clearly delineated and might overlap with other constructs without delineation, as mentioned by Hodson. The resulting fuzziness is also a possible explanation of the jingle jangle fallacy. We do recognize that overlap is perceived as an issue and would like it if Hodson was right. Unfortunately, this overlap might be a direct consequence of the fuzziness that mirrors psychological reality. We cannot force reality to behave in line with our need to categorize and to delineate. Subjective differences between and within scientists in experiencing and interpreting the psychological reality might explain that the same reality is categorized in varying ways and that different sets of indicators are selected for the same property. Such differences between subjective experiences of scientists are not what we mean by varying constructs, although it could explain some of the divergence in the psychological literature. Again, we do not claim that our view on constructs and the psychological reality is the only valid (or even plausible) one, and we would like to keep the door open for endeavors with a different premise from which to start, in tandem with those called for by Bollen, Hodson, and Machery.

As suggested by Bollen and Hodson, the fuzziness idea might encourage some authors to ignore continuing replication failures. To counter the realistic fear that scientists stick with non-replicated findings, one can use methods based on

generalization principles, ideally beginning from the early stages of testing a research hypothesis. Generalization principles are what underlies several methods discussed in our Questioning paper. Examples of such methods are meta-studies, uncertainty-based power analysis, and conceptual replications. Meta-studies (DeKay et al., 2022) used in early stages of testing a research hypothesis help to prevent non-replicable findings from being published. Power estimates based on uncertainty of effects (Pek & Park, 2019; Pek et al., 2022) can help prevent use of unrealistically large power estimates. Conceptual replications and variation of methods help to prevent the making of general claims when referring to specific non-generalizable findings. As a general perspective, we believe it is desirable to embrace uncertainty as suggested in Cronbach's (1982) paper "In praise of uncertainty," and to realize that statistics might quantify uncertainty, but does not reduce it. There will always be false positives, but we can use more realistic methods to quantify uncertainties using generalization principles. Generalizability theory as initiated by Cronbach et al. (1963) is a rather specific psychometric approach as an extension of classical test theory, whereas generalization principles constitute a broader and more generally applicable set of principles for the interpretation of findings (e.g., De Boeck & Jeon, 2018; Shrout & Rodgers, 2018).

We appreciate that Hodson seems to agree that "If pulled off successfully, this [our Questioning paper approach] would be a serious achievement and advance in the field." So why not keep that door open, allow for parallel efforts, and be optimistic instead of betting on just one horse? We also believe that the dartboard analogy as explained in Hodson's commentary and supported by papers he refers to, is part of what we suggest when discussing content validity (another reason to represent constructs as areas). The similarity Hodson noticed between his dartboard idea and our proposal is that different subsets of construct content (indicators) can be represented as different points on the dartboard. However, Hodson's call for refinement of constructs further on in his commentary paper might lead to being overly selective in the choice of indicators goodness of fit of measurement models and to improve the reliability of measures instead of trying to cover the whole content of the construct.

Looking back at the methods we see as possible implementations of our view, we acknowledge that these methods are admittedly complex (but still simpler than possibly endless interaction terms). We believe that a complex reality requires complex methods and analysis models to deal with a complex reality that can make inferences as suggested by González and Irribarra. For example, complex factor models for experimental effects are proposed by McShane et al. (2022) and by De Boeck, DeKay, and Xu (2022), which were not mentioned in our Questioning paper. These papers show how generalizable and heterogeneous effects are reflected in dimensions of a factor model.

Operational Definitions and Measurement

In Uher's commentary, operational definitions (ODs) have a prominent role. We believe that her ideas overlap with how ODs are discussed in our Questioning paper, except for her critical remarks on measurement based on ODs. Uher

claims that ODs are a necessity and continues that "Constructs are multi-referential conceptual entities—they refer to qualitatively heterogeneous referents. Scientists must decide which particular qualities are representative of a construct. 'Extraversion' may be operationalized with various behaviors, thoughts and emotions, thus qualitatively different phenomena. Joint consideration of their different qualities defines the construct's meaning." In other words, elements of an OD (such as behaviors, thoughts, and emotions) are referents of the construct. They correspond to what Cronbach and Meehl (1955) call the content of the construct. In the common psychometric terminology, they are called indicators of the construct or latent variable as the constructs' statistical representation. Uher's idea of constructs being multi-referential corresponds to the heterogeneity we assign to constructs in our Questioning paper, and our representation of constructs as areas. A set of referents for extraversion would be a set of rather specific properties (e.g., outgoing, talkative). Remember that we use the term construct for both, the property and the concept of a property that is conceptualized. One can easily see that a hierarchy with overlap can be built based on low-abstraction and high-abstraction properties. A high-abstraction property might apply to a broad set of low-abstraction properties. The concept of a high-abstraction property, for example "valence of the object evaluation" as a definition of attitude, is a definition referring to one property, which suggests there is homogeneity. In that sense, the attitude construct is homogenous, but it is also heterogenous in that lower-abstraction properties (e.g., positive opinions, positive feelings, positive approach behaviors) and their concepts are qualitatively different in Uher's view.

We, however, disagree with Uher's remark that measurement across qualitatively different properties is impossible. We understand that one might have issues with combining apples and pears, but we also believe that it depends on how measurement is defined and whether highly abstract properties can be measured. It can make sense to measure how much pome fruit one has, which would include apples and pears. In our Questioning paper, we took the perspective of a psychometrician in terms of beliefs about measurement, following from Cronbach and Meehl (1955), and common measurement practices in psychology (heavily criticized by Uher). The price we pay for measuring heterogeneous constructs comes in the form of potential violations of measurement invariance and of multiple levels of multidimensionality due to the qualitatively different low-abstraction properties. In line with the idea about fractals expressed in our Questioning paper, it seems that even low-abstraction properties can be decomposed in even lower-abstraction properties. The statistical evidence for levels of multidimensionality is discussed in our Questioning paper referring to personality. Each of the big five personality dimensions is multidimensional in terms of facets, and facets in turn are multidimensional in terms of item-specific components, aligning with Spearman's two-factor theory. This complexity is one of the motivations behind our Questioning paper. Rather than the quite extreme but logically pure conclusion of Uher that measurement is impossible if constructs are multi-referential, we have looked for methods to deal with the complexity such

as measurement continuity. Starting from the same premise as Uher (i.e., heterogeneous constructs), we come to a different conclusion.

We also agree with Uher that ODs such as self-ratings, can easily be compromised because of confounding between method and object of research. The method and the object of research overlap in that the method relies on psychological functioning (problem solving in cognitive tests and judgments in self-ratings) while the object of the research is also psychological functioning. De Boeck and Gore (2023) use the term “Janus face” for this type of confounding in psychometric approaches. A related remark Uher makes is the explanatory circularity when psychological data such as test responses are used to measure a trait and to explain the very same test responses.

Finally, we like to highlight a topic for discussion of ODs and measurement. Although we assigned an important role to the content of a construct and its OD (in the sense of content validity), we want to highlight some important issues related to the C-OD gap that remains to be discussed. According to McDonald (1999, p. 201) a construct is abstractive or existential. The distinction is important for the notions of content of a construct and its OD. For McDonald, abstractive refers to “abstractions from what common sense would regard as observable,” whereas existential refers to “postulated entities not (currently) observable.” (p. 201). In McDonald’s opinion, abstractive as a qualification applies to “most of those quantifiable attributes measured as achievements, abilities, personality traits, interests, and attitudes” (p. 201). It would be of interest to follow up on the distinction between what is abstractive versus existential, and what this distinction implies for ODs. For example, for existential constructs, one could argue that content validity is irrelevant and that different statistical models should be used. Markus and Borsboom (2013) associate existential constructs with causality between the measured property and its indicators, as in a reflective latent variable model.

The Relevance of Epistemological Theories

Focusing on the main issues discussed in our Questioning paper, our perspective on epistemology is rather pragmatic. We chose to discuss whether and how much epistemological theories would make a difference for our revisited notion of constructs and for what to do next. We tend to believe, for reasons to be explained, that the issues we raise in our Questioning paper do not need to be substantially modified as far as realist, logical positivist, or instrumentalist views are concerned. We will also explain why the same is not true for the social constructivist approach and for the critical realist perspective mentioned by Uher. As clarified earlier, we take a realist perspective by claiming that the psychological reality is complex and variable. Therefore, we will also discuss the local realist perspective mentioned by Haig.

Complexity and variability for a logical positivist would mean that the correspondence rules are complex and variable, and for an instrumentalist that predictions are complex and variable. From a constructivist approach, the main

source of complexity would be that observations are differently construed based on a context-specific interpretation of the observations. Whereas the critical realist perspective advocated by Uher accepts the reality of particulars (her words), but any abstraction and generalization regarding these particulars is based on the human mind and stem from a variety of perspectives one can take. See Uher’s reference to Bhaskar and Danermark (2006) for a more extensive discussion of critical realism. We believe that constructivist and critical realist perspectives would lead to a kind of questioning of constructs that assigns a much heavier role to construal processes as defined by Uher.

Because we have used rather realist language in line with the common belief of psychologists, we have more closely studied Mäki’s (2005) local realism, referred by Haig in his commentary. Local realism means realism of a unit of science (i.e., “particular theories or particular models and their parts and posits”, p. 232), which we understand as constructs, phenomena, and their relations. Furthermore, these units can vary in their degree of realism: “we may make realism itself a variable” (Mäki, p. 232). Local realism allows for variation among literally true, probably true, approximately true, probably approximately true (cited by Mäki, p. 236, from Stanford, 2003, p. 553), all “without violating realism” (Mäki, p. 242). Mäki believes foremost in varying degrees of realism of causal relations and seems to care less about the realism of “theoretically postulated unobservables” (p. 246), which explains why Haig is not focused on constructs. If interpreted correctly, local realism fits well with the complexity and variability we see in the psychological reality, except that we focus more on constructs and thus on the hypothesized un-observables.

We are aware that Haig’s epistemological interpretation of Cronbach and Meehl (1955) is different from ours. This difference in opinion would be worthy of a further discussion, but given our choice as explained in the beginning of this section, we do not see important reasons to disagree in our response.

Conclusion

The motivation for our Questioning paper was to offer an alternative explanation for important issues in psychology such as variable findings and complexities: violations of measurement invariance, moderate correlations between measures of the same construct, levels of multidimensionality, and other issues discussed here. Right now, any explanation and recommended action would be speculative, despite the growing consensus in one direction.

One dominant option assumes that these problematic issues stem from suboptimalities in the methodological and measurement approaches of psychological researchers, calling for refined definitions and improved measurement as a logical solution. Although striving for better definitions and measurement is a potential explanation, we would like to consider an alternative explanation that the psychological reality is complex and variable. It may require constructs that can accommodate the complexity and variability, and a

methodology to deal with larger than desirable uncertainties. This is why a generalization perspective is helpful. We have offered methodologies associated with this perspective in our Questioning paper, and we have tried to explain them further here, primarily in the section on sources of variability.

The generalization notion may lead to the acceptance of a random component in construct validity and to more accurate estimates of uncertainty. The ultimate goal is to make better generalizable inferences that build toward a more robust discipline grounded in well-founded theories. One might consider replication, generalization, and integration as three important challenges of empirical disciplines. Generalization (broadening and assessing the range of applicability) is more important than replication (a binary outcome of replicating an effect versus not), and integration seeks to bridge standalone findings (isolated pieces) to form a more complete whole (of the psychology puzzle). Therefore, we also want to remind ourselves and the reader of a topic that has been discussed toward the end of our Questioning paper (see the example on the wax layer of apples). The topic in question is multiscale research to integrate phenomena, requiring theorizing at different scale levels from low-abstraction properties to high-abstraction properties, including their relations and the processes in which they are involved.

It could turn out in the near or far future, that some of the approaches discussed in our Questioning paper, in the commentaries, or in our responses work better, or that some new approaches take over. However, it would not be wise to close the door on alternative approaches before we know where they might lead us. Right now, we do not know whether psychological constructs will eventually turn out to end up as temperature-like constructs. Perhaps they do, perhaps they do not. Finally, we hope that there will be opportunities for further dialogue after the commentary papers and our responses are made available to the community of psychological researchers.

Disclosure Statement

We have no known conflict of interest to declare.

ORCID

Paul De Boeck  <http://orcid.org/0000-0002-0884-2582>

References

- Bhaskar, R., & Danermark, B. (2006). Metatheory, interdisciplinarity and disability research: A critical realist perspective. *Scandinavian Journal of Disability Research*, 8(4), 278–297. <https://doi.org/10.1080/15017410600914329>
- Cronbach, L. J. (1982). In praise of uncertainty. *New Directions for Program Evaluation*, 1982(15), 49–58. <https://doi.org/10.1002/ev.1310>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16(2), 137–163. <https://doi.org/10.1111/J.2044-8317>
- Davids, K., & Olivier, N. (2008). English middles with mental and verbal predicates: Towards a typology. *English Text Construction*, 1(2), 169–197. <https://doi.org/10.1075/etc.1.2.02dav>
- De Boeck, P., DeKay, M., & Xu, M. (2022). The potential of factor analysis for replication, generalization, and integration. *Journal of the American Statistical Association*, 117(540), 1622–1626. <https://doi.org/10.1080/01621459.2022.2096618>
- De Boeck, P., & Gore, L. R. (2023). The Janus face of psychometrics. In A. van der Ark, Emmons, W., & Meijer, R. (Eds.), *Practical measurement essays on contemporary psychometrics* (pp. 31–46). Springer.
- De Boeck, P., & Jeon, M. (2018). Perceived crisis and reforms: Issues, explanations, and remedies. *Psychological Bulletin*, 144(7), 757–777. <https://doi.org/10.1037/bul0000154>
- DeKay, M., Rubinchik, N., Li, Z., & De Boeck, P. (2022). Accelerating psychological science with metastudies: A demonstration using the risky-choice framing effect. *Perspectives on Psychological Science*, 17(6), 1704–1736. <https://doi.org/10.1177/17456916221079611>
- Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, 108(2), 346–369. <https://doi.org/10.1037/0033-295X.108.2.346>
- Flake, J., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465. <https://doi.org/10.1177/2515245920952393>
- Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, 10(4), 371–388. <https://doi.org/10.1037/1082-989X.10.4.371>
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English. Part 1. *Journal of Linguistics*, 3(1), 37–81. <https://doi.org/10.1017/S0022226700012949>
- Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, 27(6), 930–957. <https://doi.org/10.1037/met0000303>
- International Organization for Standardization (ISO). (2022). *ISO 704: 2022, Terminology work—Principles and methods* (4th ed.). ISO.
- Mäki, U. (2005). Reglobalizing realism by going local, or (how) should our formulations of scientific realism be informed about science? *Erkenntnis*, 63(2), 231–251. <https://doi.org/10.1007/s10670-005-3227-6>
- Mari, L., Wilson, M., & Maul, A. (2019). *Measurement across the sciences*. Springer. <https://doi.org/10.1007/978-3-030-65558-7>
- Markus, K.A., & Borsboom, D. (2013). *Frontiers of test validity and theory: Measurement, causality, and meaning*. Routledge.
- McDonald, R.P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.
- McShane, B. B., Böckenholt, U., & Hansen, K. T. (2022). Variation and covariation in large-scale replication projects: An evaluation of replicability. *Journal of the American Statistical Association*, 117(540), 1605–1621. <https://doi.org/10.1080/01621459.2022.2117703>
- Pek, J., & Park, J. (2019). Complexities in power analysis: Quantifying uncertainties with a Bayesian-classical hybrid approach. *Psychological Methods*, 24(5), 590–605. <https://doi.org/10.1037/met0000208>
- Pek, J., Pitt, M., & Wegener, D. (2022). Uncertainty limits the use of power analysis. *Journal of Experimental Psychology General*. <https://doi.org/10.1037/xge0001273>
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, 69(1), 487–510. <https://doi.org/10.1146/annurev-psych-122216-011845>
- Stanford, P. K. (2003). Pirrhc victories for scientific realism. *Journal of Philosophy*, 100(11), 553–572.
- Wu, H., & Browne, M. W. (2015). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, 80(3), 571–600. <https://doi.org/10.1007/s11336-015-9451-3>