CRediT where Credit is Due:

A Comment on Leising et al. (2021)

Emorie D. Beck[1,*], Clifford I. Workman[2,*], & Alexander P. Christensen[2,*]

[1] Northwestern University Feinberg School of Medicine

[2] University of Pennsylvania Perelman School of Medicine

* All authors contributed to this work equally

Authors' Note. Author order was determined by five random shuffles of the authors' names in Excel. Correspondence concerning this article should be addressed to Emorie D. Beck, 633 N St. Clair St., Chicago, IL 60611. E.D.B. was supported by National Institute on Aging grants R01-AG067622 and R01-AG018436. C.I.W. was supported by the National Institute of Dental & Craniofacial Research grant F32DE029407. A.P.C. was supported by a grant funded by the Templeton Religion Trust. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the National Institutes of Health or the John Templeton Foundation. E.D.B., C.I.W., and A.P.C. all fulfilled the CRediT roles of Conceptualization, Investigation, Method, Project Administration, Visualization, Writing: Original Draft, and Writing: Review and Editing.

CRediT where Credit is Due:

A Comment on Leising et al. (2021)

Leising and colleagues propose a 10-step checklist that they argue will facilitate "a better personality science." Although we agree with many of the proposed steps, whether the checklist separates "good research" from bad is an empirical matter. We therefore consider whether the checklist would have caught one of the replication crisis's most infamous papers—namely, Bem's (2011) "Feeling the future" in the *Journal of Personality and Social Psychology*. Table 1 demonstrates the difficulty we faced in coming to a consensus on the score to assign Bem's paper, which is the first of our several criticisms.

Table 1. A worked example of Bem (2011) using the 10-step checklist proposed by Leising and colleagues.

#	Criteria	Max Score	CIW	EDB	AC
0	Paper gets published in a peer reviewed outlet	1.0	1.0	1.0	1.0
1a	Presents broad consensus regarding important research goals	5.0	0.0	0.0	0.0
1b	Addresses important research goals that were outlined in consensus document	0.5	0.5	0.5	0.5
2a	Presents broad consensus regarding terminology	5.0	0.0	5.0	0.0
2b	Uses terminology from consensus document	0.5	0.5	0.5	0.5
3a	Presents broad consensus regarding measurement practices	5.0	0.0	0.0	0.0
3b	Uses measurement practices from consensus document	0.5	0.5	0.5	0.5

4a	Presents broad consensus regarding data pre-processing and/or analysis	5.0	0.0	0.0	0.0
4b	Uses consensus practices regarding data pre-processing and/or analysis	0.5	0.5	0.5	0.5
5a	Presents broad consensus regarding state of knowledge and/or theory development	5.0	5.0	0.0	0.0
5b	Builds directly on consensus document regarding state of knowledge and/or theory development	0.5	0.0	0.5	0.5
6a	Includes algebraic or formal-logic formulation of theory being tested, and how it relates to measured variables	2.0	0.0	0.0	0.0
6b	Includes account of how the tested formal theory relates to previous formulations of the same or related theories	1.0	0.0	0.0	0.0
7a	Strictly separates explorative from confirmatory analyses, with the latter being pre-registered at the same level of specificity at which the results are later reported	1.0	0.0	0.0	0.0
7b	Is a registered report	2.0	0.0	0.0	0.0
8	Includes at least one direct replication attempt (of others' or one's own results), with a new sample and at least equal power as previous study	1.0	1.0	1.0	1.0
9a	Includes pre-registered a priori power analysis / sample size planning based on specific and realistic expected effect size estimate	0.5	0.5	0.5	0.5
9b	Has an expected type I error rate of <= .05 and type II error rate of <= .20, based on realistic effect size estimates	1.0	1.0	0.0	0.0
9c	Demonstrates representativeness of participant samples(s) in regard to the population of interest	3.0	0.0	1.5	1.0
9d	Demonstrates representativeness of stimuli in regard to the environmental conditions of interest	3.0	0.0	3.0	3.0

10a	Data is made open	0.5	0.5[1]	0.5	0.0
10b	Open data is accompanied by meta-data that (at least) documents all variables in the data set in a manner that enables new analyses without requiring further interactions with the people who collected the data (see FAIR principles)	1.0	0.0	0.0	1.0
10c	Code is made open (and well documented)	0.5	0.0	0.5[2]	0.0
10d	Materials are made open (and well documented)	0.5	0.0	0.0	0.0
10e	All data, materials and code from a project are found in a single directory online	0.5	0.0	0.0	0.0

The midpoint of our scores was 12.0 (building consensus: 3.3, using consensus: 2.3, formalization: 0.0, preregistration: 0.0, replication: 1.0, informativeness: 3.5, and open science: 0.8). Does this score separate the bad from the good, the reproducible from the irreproducible? These questions are difficult to answer for at least two reasons. Since most published research hasn't been scored, individual scores are difficult to contextualize. However, even if most published research was scored, and a consensus between scorers was reached, we contend that conceptual problems built into the checklist render scores difficult to interpret, and that the scope of the checklist misses important things.

Consensus Statements

A critical component of Leising and colleagues' steps toward improving scientific standards in personality center around consensus building. There are several critical ways in which the methods for building consensus in psychology could have unintended negative

[1]Available from https://replicationindex.com/2018/01/20/bemcorrespondence/
[2]Available through contact with Bem (2011)

consequences by assuming: (1) our science is sufficiently mature for consensus to emerge and (2) that consensus building will change incentives in ways that do not reward well-known, eminent, and productive individuals (in terms of publication numbers).

Eminence and Advancement

First, it is unclear whether and how early career researchers (ECRs) and researchers from underrepresented backgrounds (RUBs) will have a role in consensus building. The proposed system rewards individuals for collaborating with others to build consensus. If past initiatives and expert meetings in personality are a representative sample, however, then consensus will be driven by a small group of mid- to late-career researchers from the United States and Western Europe. In part, this arises from eminence with many eminent personality scholars having little contact with other researchers outside of western nations. Thus, how ECRs and RUBs will play a role in consensus building results in an unfair penalization of location and rank.

Second, there is an inherent tension between consensus and innovation. Given the overlap between eminent scholars with those in reviewer, editor, and other positions of influence, consensus statements are prone to enabling undue gatekeeping against challenges to consensus. Or, at minimum, publishing contradictory statements and research becomes prohibitively difficult, particularly for ECRs and RUBs, because researchers who disagree with the consensus may adhere for the sake of the proposed standards of "quality."

Finally, scholars require adequate training in the requisite domain(s) to create consensus documents in them. Yet the current academic system rewards individual contributions, particularly empirical ones, more than team-science-based or theoretical contributions. For consensus documents to guide research, personality science needs to (1) embrace team-science by rewarding many types of contributions, and (2) train students in theory building as much as

statistics.

A CRediT Alternative

Creating a better science requires a shift in academia's reward structures. The current system rewards producing more publications with little reference to contributions to those publications. Even with Leising and colleagues' ten steps toward a better science, researchers who produce more "quality" research with minimal contribution will be most rewarded. The contributor roles taxonomy (CRediT)—fourteen high-level roles that specify the researcher's contribution to a publication—offers a simple yet effective way of weighing the quality of researcher contributions rather than quantity alone. We propose adding a CRediT section to CVs to move us toward contribution-based standards (see Figure 1 for an example).

Contributorship Roles (CRediT)

Paper	Year	Concept	Data Cur.	Analysis	Funding	Invest.	Method.	Admin.	Resources	Software	Supervis.	Validation	Vis.	Draft	Review
Invited Journal Articles															
Beck & Christensen	2021	X	–	–	X	X	X	X	–				X	X	X
Jayawickreme et al.	2021	X	–	–	–	–	–		–	–		–	–	X	X
Beck & Jackson	2020a	X	X	X	X		X	X		–		X	X	X	X
Nosek et al.	2019	X	–	–	–	–			–	–			X		X
Journal Articles															
Beck & Jackson	2021a	X	X	X			X	X	X	–		X	X	X	X
Bollich-Ziegler et al.	2021	X	X	X			X	X	X	–			X	X	X
Malle et al.	2021	X		–			X		X	–					X
Beck & Jackson	2021b	X	X	X	X	X	X	X	X	–		X	X	X	X
Saef et al.	2021	X	X	X	–	–	–	X	–	–		–	X	X	X
Hill et al.	2021	X	X	X			X	X	X	–		X	X	X	X
Jackson & Beck	2021a	X	X	X	X	X	X	X	X	–			X	X	X
Jackson & Beck	2021b	X	X	X			X	X	X	–		X	X	X	X
Frumkin et al.	2020	X	X	X			X		X	–			X	X	X
Beck & Jackson	2020b	X	X	X			X	X	X	–		X	X	X	X
Piccirillo et al.	2019	X	X	–	–		–	X	X	–		–	X	X	X
Beck & Jackson	2017	X	X	X			X	X		–			X	X	X

Figure 1. Example of CRediT Section in a CV. "X" indicates contributorship role, "–" indicates not applicable for a given project, and blanks indicate roles not undertaken. "Concept" = Conceptualization, "Data Cur." = Data Curation, "Invest" = Investigation, "Method" = Methodology, "Admin" = Project Administration, "Supervis." = Supervision, "Vis." = Visualization, "Draft" = Writing: Original Draft, "Review" = Writing: Review & Editing.

Besides making a researcher's contributions clear, the CRediT section offers insights into expertise and team science. A quantitative role may include "analysis." A supervisory role may include "conceptualization" and "funding." "Writing: original draft" indicates that the researcher contributed to their content area. Different combinations of contributions reflect different yet important roles in team science.

Theory Training

The target article is the latest to join in calling for more formal theory (e.g., Borsboom et

al., 2021; Gray, 2017; Oberauer & Lewandowsky, 2019). While we agree that better theory can help move psychology forward (Maatman, 2021), psychology trainees aren't trained to think theoretically (Bosch, 2018). If psychology is to improve theory, then psychologists must be trained in theory (Smaldino, 2019). Trainees receive instruction on basic and advanced statistics, yet they often do not receive training on basic theory. Instead, students take courses that are within their area of specialization (Bosch, 2018). Learning about extant theorizing is a far cry, however, from learning how to theorize.

Borsboom and colleagues (2021) outline a theory construction course that provides an example for training programs. Students learn to distinguish between data, phenomena, and theories. Then, students choose a topic in psychology to identify robust phenomena. Finally, students use software for simulations to create models that test theoretical propositions. This program could be split into separate courses where students learn to understand the difference between modeling and theory (Haslbeck et al., 2019), simulate models to test theories (Robinaugh et al., 2020), and investigate incompatibilities and underdetermination in theory (Maatman, 2021).

Conclusion

Leising and colleagues' rubric fails to address underlying systemic issues in psychology. Consensus building, while important, will only reify eminence and gatekeeping. Rating research based on a "quality-based" checklist as opposed to a contribution-based rubric perpetuates the score counting that is endemic to the current reward system. Finally, formal theories cannot be achieved without formal training in theory. These systemic issues are pervasive and cannot be fixed without changes to the reward structure. Without changing the evaluation system of academia, we cannot change the reward system that supports it. Including the CRediT taxonomy

in CVs offers a simple yet effective way of weighing the quality of researcher contributions rather than quantity alone.

References

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*(3), 407–425. https://doi.org/10.1037/a0021524

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756–766. https://doi.org/10.1177/1745691620969647

Bosch, G. (2018). Train PhD students to be thinkers not just specialists. *Nature*, *554*, 277. http://dx.doi.org/10.1038/d41586-018-01853-1

Gray, K. (2017). How to map theory: Reliable methods are fruitless without rigorous theory. *Perspectives on Psychological Science*, *12*(5), 731–741. https://doi.org/10.1177/1745691617691949

Haslbeck, J., Ryan, O., Robinaugh, D., Waldorp, L., & Borsboom, D. (2019). Modeling psychopathology: From data models to formal theories. *PsyArXiv*. https://doi.org/10.31234/osf.io/jgm7f

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, *26*(5), 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

Maatman, F. O. (2021). Psychology's theory crisis, and why formal modelling cannot solve it. *PsyArXiv*. https://doi.org/10.31234/osf.io/puqvs

Robinaugh, D. J., Haslbeck, J. M., Ryan, O., Fried, E. I., & Waldorp, L. J. (2020). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction.

Perspectives on Psychological Science, *16*(4), 725–743.

https://doi.org/10.1177/1745691620974697

Smaldino, P. (2019). Better methods can't make up for mediocre theory. *Nature*, *575*, 9.

https://doi.org/10.1038/d41586-019-03350-5