

Preregistration

Inter-Rater Reliability in Judging the Quality of Research Articles

Anna Seyffert-Müller, Lucie Kreuzer & Daniel Leising

Technische Universität Dresden

December 7th 2022

1. Purpose of this study

In this study, we investigate inter-rater reliabilities for ratings of various desirable qualities of research articles. Establishing this is a key pre-requisite for being able to evaluate the overall quality of a body of scientific work, which may be relevant at the individual level (e.g., in making decisions over hiring, promotion and funding), as well as in ranking research institutions. The present study is basically a more highly powered attempt to replicate a pilot study that was reported on in Leising et al., (2022b; page 12). That previous study found good inter-rater reliability for an overall score of methodological rigor, $ICC(1,1) = .81$, but much lower inter-rater reliability for ratings of article content referring to consensus-building, with $ICC(1,1)$ ranging from $-.01$ to $.59$ for the respective criteria. The latter finding was attributed mainly to difficulties finding enough articles with such content. The present study will use a new, larger, and more diverse set of research articles to be rated. It will also use a somewhat revised rating scheme that is “ready-to-use” (i.e., using it does not require any additional instruction or training).

2. Endpoint and Sample Size Calculation

The primary endpoint in this study will be the level of inter-rater agreement for an overall score of methodological rigor. As in the previous study, this endpoint will be based on all the quality criteria that pertain to empirical studies (numbers 5 to 9 – please note the renumbering compared to the first version of the rating scheme). The overall score will be derived by simply *averaging* across all of these criteria. Thus, it will reflect the raters’ ability to reliably detect the presence vs. absence of various quality markers, on average.

We assume that an $ICC(1,1)$ of $.60$ or more is acceptable for ratings of this type. Using this threshold, an expected ICC of $.80$, an Alpha (two-tailed) of $.01$, and a Beta of $.10$, the required number of papers to be rated is 62 (<https://wnarifin.github.io/ssc/ssicc.html>; Arifin, 2022; Walter et al., 1998). Accounting for a possible dropout rate of 10%, we will need an overall sample of at least 69 articles. However, we decided to gather a considerably larger sample of articles ($n = 110$), in order to also enable analyses of inter-rater reliability at the level of individual criteria, with sufficient precision.

Our secondary endpoints will be the respective reliabilities for all individual criteria, including those pertaining to consensus content (1 to 4, and 10). These will be reported descriptively, including 95% confidence intervals.

We will also descriptively report ICC(1,3) for all individual criteria, and for the overall score of methodological rigor. We will report the distributions (percentiles) of ratings for all criteria, averaged across raters. And we will investigate the extent to which the individual criteria may be rated independent of one another, by comparing correlations between raters for the same criteria with correlations between raters for different criteria. This basically amounts to a multi-trait-multi-method analysis (Campbell & Fiske, 1959) in which the raters are the “methods”. The mono-trait-hetero-method correlations should be considerably higher than the hetero-trait-hetero-method correlations. This analysis requires a sample of articles in which the individual criteria are separable from one another (i.e., pairs of criteria are not always met by the same articles). We compiled our sample with that consideration in mind.

3. Article Sample

Inter-rater agreement may only be achieved to the extent that there is variation between the objects that are being rated, on the variables of interest. To ensure such variation, we (AS & LK) collected a sample of 80 relatively current articles that we thought meet one or several of the 38 criteria.

We also added a random sample of 30 relatively current articles using PubMed (15 articles) and EBSCOhost (15 articles) as databases. This latter sub-sample of articles was to represent “the typical research article published at present”, with *no* selection for research quality taking place whatsoever. All of the 110 articles will be judged in regard to the same criteria, and judges for all articles will be drawn from the same overall group of research assistants. The complete list of articles may be found in the Appendix A.

4. Raters

Ratings will be provided by research assistants who will not receive any additional training or instructions but rather use the rating scheme “as is” (the rating scheme may be found here: <https://osf.io/mbggq3/>). This is important to ensure that the inter-rater reliabilities we obtain in this study will be generalizable to any other group of relatively untrained users. It is also important to make using the rating scheme as easy as possible for everyone. Raters will be instructed to strictly avoid any communication with one another about their ratings.

Each article will be rated by three raters, with nobody rating the same article twice. Because not all raters will be able to handle the same workload, we will fill an urn with 330 lots (three for each article). Whenever a rater signals availability we will draw a lot from this urn and assign the respective article to that rater. This drawing will be performed without replacement, so once

an article has been drawn and assigned three times, there will be no more lots for it in the urn. Only if a rater would be assigned the same article twice will we put the respective lot back in the urn and draw again.

5. Rating Scheme

The rating scheme was partially revised with the aim of making it “ready-to-use” without additional training or instruction. This required adding a few explanations and clarifications to the text of some items, as compared to the previous Version 2 of the rating scheme (which may be found here: <https://osf.io/mbggq3/>). All 38 criteria of the current (third) version will be rated using the response options “does not apply”, “partly applies”, “does apply” and “unclear”. Ratings will be provided online using SoSci Survey. The current version of the rating scheme may be found in Appendix B and here (<https://osf.io/mbggq3/>). Every rater will also be asked to report the time that he or she needed for rating each article.

References

- Arifin, W. N. (2022). Sample size calculator (web). Retrieved from <http://wnarifin.github.io>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81.
- Leising, D., Thielmann, I., Glöckner, A., Gärtner, A., & Schönbrodt, F. (2022a). Ten steps toward a better personality science – a rejoinder to the comments. *Personality Science*, 3, e7961. <https://doi.org/10.5964/ps.7961>
- Leising, D., Thielmann, I., Glöckner, A., Gärtner, A., & Schönbrodt, F. (2022b). Ten steps toward a better personality science – how quality may be rewarded more in research evaluation. *Personality Science*, 3, e6029. <https://doi.org/10.5964/ps.6029>
- Walter, S.D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in medicine*, 17, 101-110.

Appendix A – List of Research Articles

- Aczel, B., Szaszi, B., Nilsonne, G., van den Akker, O. R., Albers, C. J., van Assen, M. A., Bastiaansen, J. A., Benjamin, D., Boehm, U., Botvinik-Nezer, R., Bringmann, L. F., Busch, N. A., Caruyer, E., Cataldo, A. M., Cowan, N., Delios, A., van Dongen, N. N., Donkin, C., van Doorn, J. B., ... Wagenmakers, E.-J. (2021). Consensus-based guidance for conducting and reporting multi-analyst studies. *ELife*, 10, e72185. <https://doi.org/10.7554/eLife.72185>
- Ajeen, R., Ajeen, D., Wisdom, J. P., Greene, J. A., Lepage, T., Sjoelin, C., Melvin, T., Hagan, T. E., Hunter, K. F., Peters, A., Mercer, R., & Brancu, M. (2022). The Impact of Trauma-Informed Design on Psychological Well-Being in Homeless Shelters. *Psychological Services*. Advance online publication. <https://dx.doi.org/10.1037/ser0000724>
- Aliko, S., Huang, J., Gheorghiu, F., Meliss, S., & Skipper, J. I. (2020). A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Scientific Data*, 7(1), 347. <https://doi.org/10.1038/s41597-020-00680-2>
- Astakhova, M. N., & Ho, V. T. (2022). Passionate Leaders Behaving Badly: Why Do Leaders Become Obsessively Passionate and Engage in Abusive Supervision?. *Journal of Occupational Health Psychology*. Advance online publication. <https://dx.doi.org/10.1037/ocp0000340>
- Bishop, D.V.M. (2014), Ten questions about terminology for children with unexplained language problems. *Int J Lang Commun Disord*, 49: 381-415. <https://doi.org/10.1111/1460-6984.12101>
- Bishop, D. V. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International journal of language & communication disorders*, 52(6), 671-680.
- Bosnjak, M., Fiebach, C. J., Mellor, D., Mueller, S., O'Connor, D. B., Oswald, F. L., & Sokol, R. I. (2022). A template for preregistration of quantitative research in psychology: Report of the joint psychological societies preregistration task force. *American Psychologist*, 77(4), 602.
- Bourassa, K. J., Tackman, A. M., Mehl, M. R., & Sbarra, D. A. (2019). Psychological Overinvolvement, Emotional Distress, and Daily Affect Following Marital Dissolution. *Collabra: Psychology*, 5(1), 8. <https://doi.org/10.1525/collabra.184>
- Bourjade, M., Cochet, H., Molesti, S., & Guidetti, M. (2020). Is Conceptual Diversity an Advantage for Scientific Inquiry? A Case Study on the Concept of 'Gesture' in Comparative Psychology. *Integrative Psychological and Behavioral Science*, 54(4), 805–832. <https://doi.org/10.1007/s12124-020-09516-5>
- Brandt, M. J., & Crawford, J. T. (2019). Studying a Heterogeneous Array of Target Groups Can Help Us Understand Prejudice. *Current Directions in Psychological Science*, 28(3), 292–298. <https://doi.org/10.1177/0963721419830382>
- Carbon, C.-C. (2020). The Psychology of Wearing Face Masks in Times of the COVID-19 Pandemic. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3584834>
- Carvalho, G. F., Luedtke, K., & Braun, T. (2021). Minimal important change and responsiveness of the Migraine Disability Assessment Score (MIDAS) questionnaire. *The Journal of Headache and Pain*, 22(1), 126. <https://doi.org/10.1186/s10194-021-01339-y>
- Ciranka, S., & van den Bos, W. (2019). Social Influence in Adolescent Decision-Making: A Formal Framework. *Frontiers in Psychology*, 10, 1915. <https://doi.org/10.3389/fpsyg.2019.01915>

- Conijn, A. P., Jens, S., Terwee, C. B., Breek, J. C., & Koelemay, M. J. W. (2015). Assessing the Quality of Available Patient Reported Outcome Measures for Intermittent Claudication: A Systematic Review Using the COSMIN Checklist. *European Journal of Vascular and Endovascular Surgery*, 49(3), 316–334. <https://doi.org/10.1016/j.ejvs.2014.12.002>
- Cordingley, L., & Lee, R. R. (2020). Can we implement the new research agenda for mental health? *Nature Reviews Rheumatology*, 16(4), 191–192. <https://doi.org/10.1038/s41584-020-0399-z>
- Costello, T. H., Bowes, S. M., Baldwin, M. W., Malka, A., & Tasimi, A. (2022). Revisiting the Rigidity-of-the-Right Hypothesis: A Meta-Analytic Review. *Journal of Personality and Social Psychology*. Advance online publication. <https://dx.doi.org/10.1037/pspp0000446>
- Crum II, J. E. (2021). Future Applications of Real-World Neuroimaging to Clinical Psychology. *Psychological Reports*, 124(6), 2403–2426. <https://doi.org/10.1177/0033294120926669>
- Cui, J., Lichtwarck-Aschoff, A., Olthof, M., Li, T., & Hasselman, F. (2022). From Metaphor to Computation: Constructing the Potential Landscape for Multivariate Psychological Formal Models. *Multivariate Behavioral Research*, 1–19. <https://doi.org/10.1080/00273171.2022.2119927>
- Dawel, A., Shou, Y., Smithson, M., Cherbuin, N., Banfield, M., Caelear, A. L., ... & Batterham, P. J. (2020). The effect of COVID-19 on mental health and wellbeing in a representative sample of Australian adults. *Frontiers in psychiatry*, 11, 579985.
- Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., ... & Wang, Y. (2021). The world federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder. *Neuroscience & Biobehavioral Reviews*, 128, 789-818.
- Fiagbenu, M. E., Proch, J., & Kessler, T. (2021). Of deadly beans and risky stocks: Political ideology and attitude formation via exploration depend on the nature of the attitude stimuli. *British Journal of Psychology*, 112(1), 342-357.
- Fiedler, J., Seiferth, C., Eckert, T., Woll, A., & Wunsch, K. (2022). A just-in-time adaptive intervention to enhance physical activity in the SMARTFAMILY2.0 trial. *Sport, Exercise, and Performance Psychology*. <https://doi.org/10.1037/spy0000311>
- Finsterer, J. (2020). The metabolic hypothesis is more likely than the epileptogenic hypothesis to explain stroke-like lesions. *Wellcome Open Research*, 5.
- Fischer, R., Karl, J. A., Luczak–Roesch, M., Fetvadjev, V. H., & Grener, A. (2020). Tracing Personality Structure in Narratives: A Computational Bottom–Up Approach to Unpack Writers, Characters, and Personality in Historical Context. *European Journal of Personality*, 34(5), 917–943. <https://doi.org/10.1002/per.2270>
- Frewen, P., Oldrieve, P., & Law, K. (2022). Teaching psychology in virtual reality. *Scholarship of Teaching and Learning in Psychology*. <https://doi.org/10.1037/stl0000341>
- Fuermaier, A. B., Fricke, J. A., de Vries, S. M., Tucha, L., & Tucha, O. (2019). Neuropsychological assessment of adults with ADHD: A Delphi consensus study. *Applied Neuropsychology: Adult*, 26:4, 340-354.
- Genzel L, Dragoi G, Frank L, Ganguly K, de la Prida L, Pfeiffer B, Robertson E. (2020). A consensus statement: defining terms for reactivation analysis. *Phil. Trans. R. Soc. B* 375: 20200001. <http://dx.doi.org/10.1098/rstb.2020.0001>
- Gimenez-Serrano, S., Garcia, F., & Garcia, O. F. (2022). Parenting styles and its relations with

- personal and social adjustment beyond adolescence: Is the current evidence enough? *European Journal of Developmental Psychology*, 19(5), 749–769. <https://doi.org/10.1080/17405629.2021.1952863>
- Graham, E. K., Weston, S. J., Gerstorf, D., Yoneda, T. B., Booth, T., Beam, C. R., Petkus, A. J., Drewelies, J., Hall, A. N., Bastarache, E. D., Estabrook, R., Katz, M. J., Turiano, N. A., Lindenberger, U., Smith, J., Wagner, G. G., Pedersen, N. L., Allemand, M., Spiro, A., ... Mroczek, D. K. (2020). Trajectories of Big Five Personality Traits: A Coordinated Analysis of 16 Longitudinal Samples. *European Journal of Personality*, 34(3), 301–321. <https://doi.org/10.1002/per.2259>
- Gries, T., Müller, V., & Jost, J. T. (2022). The market for belief systems: A formal model of ideological choice. *Psychological Inquiry*, 33(2), 65-83.
- Grosz, M. P., Schwartz, S. H., & Lechner, C. M. (2021). The longitudinal interplay between personal values and subjective well-being: A registered report. *European Journal of Personality*, 35(6), 881-897.
- Hallford, D. J., Hardgrove, S., Sanam, M., Oliveira, S., Pilon, M., & Duran, T. (2022). Remembering for resilience: Brief cognitive-remembrance therapy improves psychological resources and mental well-being in young adults. *Applied Psychology: Health and Well-Being*, 14(3), 1004–1021. <https://doi.org/10.1111/aphw.12364>
- Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., ... & Frank, M. C. (2021). Analytic reproducibility in articles receiving open data badges at the journal Psychological Science: an observational study. *Royal Society open science*, 8(1), 201494.
- He, J., Zhao, Y., Zhang, H., & Lin, Z. (2021). Orthorexia nervosa is associated with positive body image and life satisfaction in Chinese elderly: Evidence for a positive psychology perspective. *International Journal of Eating Disorders*, 54(2), 212–221. <https://doi.org/10.1002/eat.23400>
- Howes, O. D., Thase, M. E., & Pillinger, T. (2022). Treatment resistance in psychiatry: state of the art and new directions. *Molecular Psychiatry*, 27(1), 58-72.
- Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, 3(2), 166-184.
- Hyatt, C. S., Sharpe, B. M., Owens, M. M., Listyg, B. S., Carter, N. T., Lynam, D. R., & Miller, J. D. (2022). Searching high and low for meaningful and replicable morphometric correlates of personality. *Journal of Personality and Social Psychology*, 123(2), 463–480. <https://doi.org/10.1037/pspp0000402>
- Kang, H., Miksche, M. S., & Ellingsen, D.-M. (2022). *personality traits and placebo effects: A preregistered systematic review and meta-analysis*. 25.
- Kim, H., Di Domenico, S. I., & Connelly, B. S. (2019). Self–Other Agreement in Personality Reports: A Meta-Analytic Comparison of Self- and Informant-Report Means. *Psychological Science*, 30(1), 129–138. <https://doi.org/10.1177/0956797618810000>
- King, P. R., Beehler, G. P., VanTreese, K., Johnson, E. M., Buchholz, L. J., & Wray, L. O. (2022). Patient acceptability, use, and recommendations to improve interventions delivered in primary care behavioral health. *Professional Psychology: Research and Practice*. <https://doi.org/10.1037/pro0000493>
- Kofler, M. J., Wells, E. L., Singh, L. J., Soto, E. F., Irwin, L. N., Groves, N. B., Chan, E. S. M., Miller, C. E., Richmond, K. P., Schatschneider, C., & Lonigan, C. J. (2020). A randomized controlled trial of central executive training (CET) versus inhibitory control training (ICT) for ADHD. *Journal of Consulting and Clinical Psychology*, 88(8), 738–756. <https://doi.org/10.1037/ccp0000550>

- Kritzler, S., Krasko, J., & Luhmann, M. (2020). Inside the happy personality: Personality states, situation experience, and state affect mediate the relation between personality and affect. *Journal of Research in Personality*, 85, 103929. <https://doi.org/10.1016/j.jrp.2020.103929>
- Krpan, D. (2020). Unburdening the Shoulders of Giants: A Quest for Disconnected Academic Psychology. *Perspectives on Psychological Science*, 15(4), 1042–1053. <https://doi.org/10.1177/1745691620904775>
- Kuper, N., Kroencke, L., Harari, G. M., & Denissen, J. J. A. (2022). Who benefits from which activity? On the relations between personality traits, leisure activities, and well-being. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspp0000438>
- Laurent, R., Barnaud, M.-L., Schwartz, J.-L., Bessière, P., & Diard, J. (2017). The complementary roles of auditory and motor information evaluated in a Bayesian perceptuo-motor model of speech perception. *Psychological Review*, 124(5), 572–602. <https://doi.org/10.1037/rev0000069>
- Leckelt, M., Richter, D., Wetzel, E., & Back, M. D. (2019). Longitudinal Associations of Narcissism with Interpersonal, Intrapersonal, and Institutional Outcomes: An Investigation Using a Representative Sample of the German Population. *Collabra: Psychology*, 5(1), 26. <https://doi.org/10.1525/collabra.248>
- Lelleck, V. V., Schulz, F., Witt, O., Kühn, G., Klein, D., Gendolla, A., Evers, S., Gaul, C., Thaçi, D., Sina, C., & Schröder, T. (2022). A Digital Therapeutic Allowing a Personalized Low-Glycemic Nutrition for the Prophylaxis of Migraine: Real World Data from Two Prospective Studies. *Nutrients*, 14(14), 2927. <https://doi.org/10.3390/nu14142927>
- Li, Y. I., Starr, L. R., & Wray-Lake, L. (2018). Insomnia mediates the longitudinal relationship between anxiety and depressive symptoms in a nationally representative sample of adolescents. *Depression and anxiety*, 35(6), 583-591.
- Lianov, L. S., Barron, G. C., Fredrickson, B. L., Hashmi, S., Klemes, A., Krishnaswami, J., Lee, J., Le Pertel, N., Matthews, J. A., Millstein, R. A., Phillips, E. M., Sannidhi, D., Purpur de Vries, P., Wallace, A., & Winter, S. J. (2020). Positive psychology in health care: Defining key stakeholders and their roles. *Translational Behavioral Medicine*, 10(3), 637–647. <https://doi.org/10.1093/tbm/ibz150>
- Lin, Y., Brough, R. E., Tay, A., Jackson, J. J., & Braver, T. S. (2022). Working memory capacity preferentially enhances implementation of proactive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0001195>
- Lippi, A., Rossi, S., & Soana, M. G. (2022). Status quo bias and risk tolerance in asset allocation decision-making. *Journal of Neuroscience, Psychology, and Economics*, 15(4), 195–209. <https://doi.org/10.1037/npe0000166>
- Luke, D. M., & Gawronski, B. (2022). Big Five Personality Traits and Moral-Dilemma Judgments: Two Preregistered Studies using the CNI Model. *Journal of Research in Personality*, 101, 104297. <https://doi.org/10.1016/j.jrp.2022.104297>
- Mac Giolla, E., & Kajonius, P. J. (2019). Sex differences in personality are larger in gender equal countries: Replicating and extending a surprising finding. *International Journal of Psychology*, 54(6), 705–711. <https://doi.org/10.1002/ijop.12529>
- Mayer, C.-H., & Vanderheiden, E. (2020). Contemporary positive psychology perspectives and future directions. *International Review of Psychiatry*, 32(7–8), 537–541. <https://doi.org/10.1080/09540261.2020.1813091>
- McAllister-Williams, R. H., Arango, C., Blier, P., Demyttenaere, K., Falkai, P., Gorwood, P., ... & Rush, A. J. (2020). The identification, assessment and management of difficult-to-treat depression: an international consensus statement. *Journal of affective disorders*, 267, 264-282.

- Meagher, B. R. (2020). Ecologizing Social Psychology: The Physical Environment as a Necessary Constituent of Social Processes. *Personality and Social Psychology Review*, 24(1), 3–23. <https://doi.org/10.1177/1088868319845938>
- Milek, A., Butler, E. A., Tackman, A. M., Kaplan, D. M., Raison, C. L., Sbarra, D. A., Vazire, S., & Mehl, M. R. (2018). “Eavesdropping on Happiness” Revisited: A Pooled, Multisample Replication of the Association Between Life Satisfaction and Observed Daily Conversation Quantity and Quality. *Psychological Science*, 29(9), 1451–1462. <https://doi.org/10.1177/0956797618774252>
- Miller, T. J. (2022). Assessing the Desire to Change Personality across Methods. *Journal of Personality Assessment*, 104(4), 447–457. <https://doi.org/10.1080/00223891.2021.1955695>
- Möttus, R., Sinick, J., Terracciano, A., Hrebícková, M., Kandler, C., Ando, J., Mortensen, E. L., Colodro-Conde, L., & Jang, K. L. (2018). Personality Characteristics Below Facets: A Replication and Meta-Analysis of Cross-Rater Agreement, Rank-Order Stability, Heritability, and Utility of Personality Nuances. *Journal of Personality and Social Psychology*. Advance online publication. <http://dx.doi.org/10.1037/pspp0000202>
- Napier, T. R., Howell, K. H., Scheid, C. R., & Miller-Graff, L. E. (2022). Empirically derived patterns of adverse childhood experiences and associations with prenatal social networks. *Psychology of Violence*. <https://doi.org/10.1037/vio0000453>
- Near, J., Harris, A. D., Juchem, C., Kreis, R., Marjańska, M., Öz, G., Slotboom, J., Wilson, M., & Gasparovic, C. (2021). Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: Experts’ consensus recommendations. *NMR in Biomedicine*, 34(5). <https://doi.org/10.1002/nbm.4257>
- Nett, N., Nett, T., Englert, J., & Gaschler, R. (2022). Think scientists - Think male: Science and leadership are still more strongly associated with men than with women in Germany. *Journal of Applied Social Psychology*, 52(8), 643-659.
- Newham, J. (2014). Complementary therapies in pregnancy: A means to reduce ill health and improve well-being? *Journal of Reproductive and Infant Psychology*, 32(3), 211–213. <https://doi.org/10.1080/02646838.2014.924228>
- Ng, N. L., Luke, D. M., & Gawronski, B. (2022). Prediction of dishonest behavior using the CNI model of moral dilemma judgment: a registered report. *Comprehensive Results in Social Psychology*, 1-28.
- Norgaard, M., Matheson, G. J., Hansen, H. D., Thomas, A., Searle, G., Rizzo, G., Veronese, M., Giacomel, A., Yaqub, M., Tonietto, M., Funck, T., Gillman, A., Boniface, H., Routier, A., Dalenberg, J. R., Betthausen, T., Feingold, F., Markiewicz, C. J., Gorgolewski, K. J., ... Ganz, M. (2022). PET-BIDS, an extension to the brain imaging data structure for positron emission tomography. *Scientific Data*, 9(1), 65. <https://doi.org/10.1038/s41597-022-01164-1>
- Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science*, 3(2), 229-237.
- Olsson-Collentine, A., Van Assen, M. A., & Hartgerink, C. H. (2019). The prevalence of marginally significant results in psychology over time. *Psychological science*, 30(4), 576-586.
- Orben, A., & Przybylski, A. K. (2019). Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. *Psychological Science*, 30(5), 682–696. <https://doi.org/10.1177/0956797619830329>
- Osth, A. F., Jansson, A., Dennis, S., & Heathcote, A. (2018). Modeling the dynamics of

- recognition memory testing with an integrated model of retrieval and decision making. *Cognitive Psychology*, 104, 106–142. <https://doi.org/10.1016/j.cogpsych.2018.04.002>
- Paul, K., Short, C. A., Beauducel, A., Carsten, H. P., Härpfer, K., Hennig, J., Hewig, J., Hildebrandt, A., Kührt, C., Mueller, E. M., Munk, A., Osinsky, R., Porth, E., Riesel, A., Rodrigues, J., Scheffel, C., Stahl, J., Strobel, A., & Wacker, J. (2022). The methodology and dataset of the coscience eeg-personality project – a large-scale, multi-laboratory project grounded in cooperative forking paths analysis. *Personality Science*, 3, e7177. <https://doi.org/10.5964/ps.7177>
- Prevolnik Rupel, V., Jagger, B., Fialho, L. S., Chadderton, L. M., Gintner, T., Arntz, A., ... & Crawford, M. J. (2021). Standard set of patient-reported outcomes for personality disorder. *Quality of Life Research*, 30(12), 3485-3500.
- Prins, J. B., Deuning-Smit, E., & Custers, J. A. E. (2022). Interventions addressing fear of cancer recurrence: Challenges and future perspectives. *Current Opinion in Oncology*, 34(4), 279–284. <https://doi.org/10.1097/CCO.0000000000000837>
- Rahnev, D., Balsdon, T., Charles, L., De Gardelle, V., Denison, R., Desender, K., ... & Zylberberg, A. (2022). Consensus goals in the field of visual metacognition. *Perspectives on Psychological Science*, 17(6), 1746-1765.
- Reardon, C. L., Hainline, B., Aron, C. M., Baron, D., Baum, A. L., Bindra, A., ... & Engebretsen, L. (2019). Mental health in elite athletes: International Olympic Committee consensus statement (2019). *British journal of sports medicine*, 53(11), 667-699.
- Robinaugh, D., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., ... & Borsboom, D. (2019). Advancing the network theory of mental disorders: A computational model of panic disorder.
- Rollo, I., Carter, J. M., Close, G. L., Yangüas, J., Gomez-Diaz, A., Medina Leal, D., Duda, J. L., Holohan, D., Erith, S. J., & Podlog, L. (2021). Role of sports psychology and sports nutrition in return to play from musculoskeletal injuries in professional soccer: An interdisciplinary approach. *European Journal of Sport Science*, 21(7), 1054–1063. <https://doi.org/10.1080/17461391.2020.1792558>
- Roos, J. M., & Kazemi, A. (2022). The five factor model of personality as predictor of online shopping: Analyzing data from a large representative sample of Swedish internet users. *Cogent Psychology*, 9(1), 2024640.
- Russell, N., & Gregory, R. (2015). The Milgram-Holocaust linkage: Challenging the present consensus. *State Crime Journal*, 4(2), 128-153.
- Salovich, N. A., & Rapp, D. N. (2022). How susceptible are you? Using feedback and monitoring to reduce the influence of false information. *Journal of Applied Research in Memory and Cognition*. <https://doi.org/10.1037/mac0000074>
- Sanabria-Mazo, J. P., Forero, C. G., Cristobal-Narváez, P., Suso-Ribera, C., García-Palacios, A., Colomer-Carbonell, A., Pérez-Aranda, A., Andrés-Rodríguez, L., McCracken, L. M., D'Amico, F., Estivill-Rodríguez, P., Carreras-Marcos, B., Montes-Pérez, A., Comps-Vicente, O., Esteve, M., Grasa, M., Rosa, A., Cuesta-Vargas, A. I., Maes, M., ... Luciano, J. V. (2020). Efficacy, cost-utility and physiological effects of Acceptance and Commitment Therapy (ACT) and Behavioural Activation Treatment for Depression (BATD) in patients with chronic low back pain and depression: Study protocol of a randomised, controlled trial including mobile-technology-based ecological momentary assessment (IMPACT study). *BMJ Open*, 10(7), e038107. <https://doi.org/10.1136/bmjopen-2020-038107>
- Sánchez, Amber M., Christopher W. Coleman, and Alison Ledgerwood. "Does temporal distance influence abstraction? A large pre-registered experiment." *Social cognition* 39.3 (2021).

- Sassenberg, K., & Ditrich, L. (2019). Research in Social Psychology Changed Between 2011 and 2016: Larger Sample Sizes, More Self-Report Measures, and More Online Studies. *Advances in Methods and Practices in Psychological Science*, 2(2), 107–114. <https://doi.org/10.1177/2515245919838781>
- Schwartz, F., Djeriouat, H., & Trémolière, B. (2021). The association between personality traits and third-party moral judgment: A preregistered study. *Acta psychologica*, 219, 103392.
- Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174. <https://doi.org/10.1016/j.im.2019.103174>
- Shapiro, M. O., Houtsma, C., Schafer, K. M., True, G., Miller, L., & Anestis, M. (2022). Moral injury and suicidal ideation among female national guard members: Indirect effects of perceived burdensomeness and thwarted belongingness. *Traumatology*. <https://doi.org/10.1037/trm0000424>
- Sharma, A., Minh Duc, N. T., Luu Lam Thang, T., Nam, N. H., Ng, S. J., Abbas, K. S., ... & Karamouzian, M. (2021). A consensus-based checklist for reporting of survey studies (CROSS). *Journal of general internal medicine*, 36(10), 3179-3187.
- Siegel, K. R., Mobley, T. P., & Sanderson, C. A. (2022). Addressing the college mental health crisis: Training students to become effective bystanders. *Psychological Services*. <https://doi.org/10.1037/ser0000720>
- Smith, B., & McGannon, K. R. (2018). Developing rigor in qualitative research: Problems and opportunities within sport and exercise psychology. *International Review of Sport and Exercise Psychology*, 11(1), 101–121. <https://doi.org/10.1080/1750984X.2017.1317357>
- Somani, A., Kar, S., & Parida, J. (2016). Depression in systemic sclerosis: Review of the neuro immunologic link and pharmacological management. *Medical Journal of Dr. D.Y. Patil University*, 9(3), 294. <https://doi.org/10.4103/0975-2870.167993>
- Stachl, C., Hilbert, S., Au, J., Buschek, D., De Luca, A., Bischl, B., Hussmann, H., & Bühner, M. (2017). Personality Traits Predict Smartphone Usage. *European Journal of Personality*, 31(6), 701–722. <https://doi.org/10.1002/per.2113>
- Stambulova, N. B., Ryba, T. V., & Henriksen, K. (2021). Career development and transitions of athletes: The International Society of Sport Psychology Position Stand Revisited. *International Journal of Sport and Exercise Psychology*, 19(4), 524–550. <https://doi.org/10.1080/1612197X.2020.1737836>
- Stapinski, L. A., Prior, K., Newton, N. C., Biswas, R. K., Kelly, E., Deady, M., Lees, B., Teesson, M., & Baillie, A. J. (2021). Are we making Inroads? A randomized controlled trial of a psychologist-supported, web-based, cognitive behavioral therapy intervention to reduce anxiety and hazardous alcohol use among emerging adults. *EClinicalMedicine*, 39, 101048. <https://doi.org/10.1016/j.eclinm.2021.101048>
- Steiner, M. D., & Frey, R. (2021). Representative design in psychological assessment: A case study using the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: General*.
- Stricker, J., Chasiotis, A., Kerwer, M., & Günther, A. (2020). Scientific abstracts and plain language summaries in psychology: A comparison based on readability indices. *PLOS ONE*, 15(4), e0231160. <https://doi.org/10.1371/journal.pone.0231160>
- Surov, I. A. (2022). Quantum core affect. Color-emotion structure of semantic atom. *Frontiers in Psychology*, 13, 838029. <https://doi.org/10.3389/fpsyg.2022.838029>
- Tal, S., Bar-Kalifa, E., Kleinbub, J. R., Leibovich, L., Deres-Cohen, K., & Zilcha-Mano, S.

- (2022). A multimodal case study utilizing physiological synchrony as indicator of context in which motion synchrony is associated with the working alliance. *Psychotherapy*. <https://doi.org/10.1037/pst0000465>
- Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, 76(1), 116–129. <https://doi.org/10.1037/amp0000622>
- Tremblay, M.S., Aubert, S., Barnes, J.D., Saunders, T. J., Carson, V., Latimer-Cheung, A. E., Chastin, S. F. M., Altenburg, T. M., Chinapaw, M. J. M., (2017). Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project process and outcome. *Int J Behav Nutr Phys Act* 14, 75 (2017). <https://doi.org/10.1186/s12966-017-0525-8>
- Van der Feltz-Cornelis, C. M., Elfeddali, I., Werneke, U., Malt, U. F., Van den Bergh, O., Schaefert, R., Kop, W. J., Lobo, A., Sharpe, M., Söllner, W., & Löwe, B. (2018). A European Research Agenda for Somatic Symptom Disorders, Bodily Distress Disorders, and Functional Disorders: Results of an Estimate-Talk-Estimate Delphi Expert Study. *Frontiers in Psychiatry*, 9, 151. <https://doi.org/10.3389/fpsy.2018.00151>
- Wang, L., Hsiao, J. H., Chan, A. B., Cheung, J., Hung, S., & Au, T. K. (2022). On becoming socially anxious: Toddlers' attention bias to fearful faces. *Developmental Psychology*. <https://doi.org/10.1037/dev0001472>
- West, R., Godinho, C. A., Bohlen, L. C., Carey, R. N., Hastings, J., Lefevre, C. E., & Michie, S. (2019). Development of a formal system for representing behaviour-change theories. *Nature Human Behaviour*, 3(5), 526–536. <https://doi.org/10.1038/s41562-019-0561-2>
- Weston, S. J., & Jackson, J. J. (2018). The role of vigilance in the relationship between neuroticism and health: A registered report. *Journal of Research in Personality*, 73, 27–34.
- White, S. W., Xia, M., & Edwards, G. (2021). Race, gender, and scholarly impact: Disparities for women and faculty of color in clinical psychology. *Journal of Clinical Psychology*, 77(1), 78–89. <https://doi.org/10.1002/jclp.23029>
- Wong, E., Heuschkel, R., Lindsay, C., Benson, S., & Zilbauer, M. (2021). The growing gap between demand and availability of clinical psychology in Paediatric Gastroenterology: A retrospective analysis of clinical routine care. *European Journal of Pediatrics*, 180(4), 1307–1312. <https://doi.org/10.1007/s00431-020-03825-y>
- Wood, W., Alm, K., Benjamin, J., Thomas, L., Anderson, D., Pohl, L., & Kane, M. (2021). Optimal terminology for services in the United States that incorporate horses to benefit people: A consensus document. *The Journal of Alternative and Complementary Medicine*, 27(1), 88–95.
- Wright, A. J., Chávez, L., Edelstein, B. A., Grus, C. L., Krishnamurthy, R., Lieb, R., Mihura, J. L., Pincus, A. L., & Wilson, M. (2021). Education and training guidelines for psychological assessment in health service psychology. *American Psychologist*, 76(5), 794–801. <https://doi.org/10.1037/amp0000742>
- Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, 45, e1. <https://doi.org/10.1017/S0140525X20001685>
- Zale, A., Lasecke, M., Baeza-Hernandez, K., Testerman, A., Aghakhani, S., Muñoz, R. F., & Bunge, E. L. (2021). Technology and psychotherapeutic interventions: Bibliometric analysis of the past four decades. *Internet interventions*, 25, 100425.
- Zhang, X.-Q., Zhang, B.-S., & Wang, M.-D. (2020). Application of a classroom-based positive psychology education course for Chinese medical students to increase their psychological well-being: A pilot study. *BMC Medical Education*, 20(1), 323. <https://doi.org/10.1186/s12909-020-02232-z>

Zhao, W. J., Richie, R., & Bhatia, S. (2022). Process and content in decisions from memory. *Psychological Review*, 129(1), 73.

Appendix B – “Ready to use” Rating Form

Criterion	Content
	<p>General Instruction: Try to be as objective as possible. That is, judge the applicability of each criterion separately, and to the best of your ability. Most published research articles will only meet a few criteria, if any. Sometimes the nature of a paper results in non-applicability of some criteria (e.g., sample size planning for a solely theoretical paper). In that case the criterion should be rated with "does not apply".</p> <p>Some criteria (1, 2, 3, 4 and 10) refer to so-called "consensus". Note that this means explicit attempts at documenting what many researchers in the respective field agree on. Consensus of this kind may be achieved by means of structured group processes (e.g., polling) which should be described in some detail in the respective paper, including the (groups of) people that were involved. Articles using or challenging previously established consensus should refer to a previous article (a "consensus document") describing the consensus and how it was achieved.</p>
0	Paper gets published in a peer reviewed journal.
1a	Documents explicit consensus regarding important research goals.
1b	Addresses important research goals that were outlined in a consensus document.
1c	Provides empirical / theoretical results that fundamentally challenge documented consensus regarding research goals.
2a	Documents explicit consensus regarding terminology.
2b	Uses terminology from a consensus document.
2c	Provides empirical / theoretical results that fundamentally challenge documented consensus regarding terminology.
3a	Documents explicit consensus regarding measurement practices.
3b	Includes standard measurement practices from a type 3a consensus document. Note that this does not preclude the additional use of other measures in the same study.
3c	Provides empirical/theoretical results that fundamentally challenge documented consensus regarding measurement practices.
4a	Documents explicit consensus regarding data pre-processing and/or analysis.
4b	Uses consensus practices regarding data pre-processing and/or analysis.
4c	Provides empirical/theoretical results that fundamentally challenge documented consensus regarding data pre-processing / analysis.
5a	Specifies a theory in a mathematical or formal-logic manner (as opposed to a "narrative" theory that is expressed only in natural language terms). By "formal-logic", we mean explications of IF-THEN relationships.
5b	Includes an account of how a newly specified formal theory relates to previous formulations of the same or related theories (theory integration).
5c	Includes a full account of how the measured variables used in a theory test relate to the parameters of the tested formal model.

6a	Explicitly distinguishes explorative from confirmatory analyses, with the latter having been pre-registered at the same level of specificity at which the results are reported. Judging this requires a comparison of the article with the actual pre-registration(s).
6b	All of the tested hypotheses were pre-registered.
6c	All operationalizations of relevant theoretical concepts were pre-registered (i.e., what will be measured how).
6d	All statistical procedures used in hypothesis tests were pre-registered.
6e	Article contains information on exact timelines as to when pre-registrations took place and when data was collected and analysed.
6f	All relevant deviations from the pre-registration are made explicit.
6g	Is a registered report (i.e., received "in principle acceptance" from an academic journal based on the planned research design, before the data was collected).
7a	Includes at least one direct replication attempt (of others' or one's own results), with a new sample. This includes "split sample" studies in which at least one part of the data ("hold-out sample(s)") is deliberately set aside for later replication attempts.
7b	The replication attempt had at least the same statistical power as the original study that it refers to.
7c	The replication attempt was explicitly pre-registered as a replication attempt.
8a	Includes a pre-registered a priori power analysis / sample size planning based on specific and realistic estimate of expected effect size(s).
8b	Has an expected type I error rate of $\leq .05$ and type II error rate of $\leq .20$, based on realistic effect size estimates.
8c	Demonstrates representativeness of participant samples(s) in regard to the population of interest (e.g., by using a random sampling method).
8d	Demonstrates representativeness of stimuli in regard to the environmental conditions of interest (i.e., the type and range of stimuli that are presented in the study are shown to resemble those of the relevant variable in real life).
9a	Data is made openly available. Judging this requires checking whether a link provided in the article to an online repository actually works.
9b	Open data is accompanied by meta-data that (at least) documents all variables in the dataset in a manner that enables new analyses without requiring further interactions with the people who collected the data.
9c	Analysis code (e.g., an R-script or SPSS syntax) is made openly available. Judging this requires checking whether a link provided in the article to an online repository actually works.
9d	Materials (e.g., questionnaire items) are made openly available. Judging this requires checking whether a link provided in the article to an online repository actually works.
9e	All data, materials and code from a project are found in the same "place" online (e.g., a project folder).

10a	Documents explicit consensus regarding the state of knowledge and/or theory development in a research area.
10b	Builds directly on a consensus document regarding the state of knowledge and/or theory development in a research area.
10c	Provides empirical / theoretical results that fundamentally challenge a documented consensus regarding the state of knowledge and/ or theory development in a research area.