

Editorial

Focus on new research topics in Psychological Test and Assessment Modeling

Klaus D. Kubinger (editor in chief)¹

Preamble: In the last years, “Psychological Test and Assessment Modeling” (formerly: Psychology Science [Quarterly]) has succeeded in establishing a journal that deals with psychology-specific statistical methods and problems, with psychometric forthcoming, and with psychological assessment advancements. This occurred despite the fact that the journal is still not included in the index by Thomson-Reuters – but consider the current trend of the self-evaluated (2-year) impact factor according to Kubinger, Heuberger, & Poinstingl (2010); 2014: 0.354, 2015: 0.370, 2016: 0.400, 2017: 0.798, 2018: 0.583. The number of citations of included papers proves the journal’s reception. According to Google Scholar, the perhaps most cited one is that of Rasch and Guiard (2004), with 186 citations; it deals with the robustness of the two-sample *t*-test. A further considerably often cited paper, i.e. 59 times, is the paper by Kubinger (2008), which points out the usefulness of the Rasch-model-based LLTM for measuring item administration effects; the paper of Rupp (2013) is, in turn, cited 45 times, and deals with person-fit research for IRT-models. Furthermore, there is the paper by Etzel, Rohrmann, and Brandt (2014), which has been cited 23 times; it delivers an inventive validation study of the State-Trait-Anger-Expression-Inventory.

Nevertheless, further effort is required in order to establish this journal. For this, in the first instant, we now request the authors more than ever to commit themselves to making the following available on demand: a) the data, b) the specification of the used software (version number and applied options included), as well as c) the applied source code if no pertinent software is used – this is due to the standards of research reproducibility established by Hothorn and Leisch (2011), and will enhance the journal’s image. In the second instant, we review in the following the last special topics/issues the journal dealt with through invited guest editors; such a review indicates which topics might be of general interest to the scientific community in the future. In the third instant, we offer a

¹ *Correspondence concerning this article should be addressed to:* Klaus D. Kubinger, c/o Division for Assessment and Applied Psychometrics, Faculty of Psychology, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria; email: klaus.kubinger@univie.ac.at

suggestion of some specific research areas that should be focused on due to lack of findings. Finally, we once again give indications on how to manage research work in order to contribute at a very high methodical standard.

To be announced: We face the challenge of processing manuscript submissions more quickly and publishing papers more promptly, ideally three to five months after submission.²

Review of the last special topics/issues

There are several special topics in “Psychological Test and Assessment Modeling”, which either summarize the actual state of knowledge in a specific area or try to give inspiration to a new research field. The list comprises: *Dispositions towards ridicule and being laughed at: Current research on gelotophobia, gelotophilia, and katagelasticism* (Proyer & Ruch, 2010), *New approaches to the study of self-regulated learning* (Stoeger & Ziegler, 2011), *Caregivers’ evaluations of young children’s development and behavior* (Kastner-Koller & Deimann, 2011), *Methodological advances in psychological and educational testing* (von Davier, 2011), *Configural Frequency Analysis (CFA) and other non-parametrical statistical methods* (Stemmler & von Eye, 2012), *Current issues in educational and psychological measurement: design, calibration, and adaptive testing* (Frey & Kröhne, 2012; Kröhne & Frey, 2013), *Current perspectives on the assessment of giftedness* (Vialle, 2013a, b), *Identifying effective learning environments* (Schorer, Baker, Stoeger, & Ziegler, 2015; Stoeger, Schorer, Baker, & Ziegler, 2015), *Advances in Rasch modeling: New applications and directions* (Brandt, Moulton, & Duckor, 2015), *Missing values in large-scale assessment studies* (Pohl & Abmann, 2015), *Measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) short forms* (Reeve & Teresi, 2016), *Current methodological issues in educational large-scale assessments* (Stadler, Greiff, & Krolak-Schwerdt, 2016, 2017), *Rater effects: Advances in item response modeling of human ratings* (Eckes, 2017, 2018), *Advances in Educational Measurement* (Frey, König, & Spoden, 2018a, b), *New trends in gifted identification* (Ziegler, Alghawi, & Reutlinger, 2019). In addition to these, the following are in progress: *Tree based methods for regression and classifications – statistical methods at the interface of graphics and statistics* (Stemmler, & von Eye, 2019, in prep.), *Establishing comparability and measurement invariance in large-scale assessments* (Khorramdel-Ameri, Pokropek, & van Rijn, 2019, in prep.), *Ambulant assessment* (Pawlik, 2020, in prep.).

From the journal’s point of view, the following topics would be attractive in the future: *The importance of using a testing sample set in addition to a training sample set within multi-dimensional prediction models; Elaborating statistical sequential testing approaches; The merits of adaptive testing; Determining a test-taker’s testability; Ad-*

² We have slightly changed the editorial board: Alison Ying Chen had to leave the board, as she became the editor in chief in another journal. We are also lucky to win Trevor Bond as a new, famous board member.

vantages and problems of applying self-assessments; Advantages of objective personality tests sensu R. B. Cattell over personality questionnaires.

Specific research to take into consideration

Psychological assessment instruments often suffer from high quality standardization. The reason is that random sampling – which makes representativity most likely – is difficult to realize. Instead, the situation arises that samples with persons are used, who are easily at a test author's disposal and give their consent to be tested. That is, in most cases they are volunteers. However, volunteers may differ essentially from the aimed-for population with respect to achievement motivation and ability, as well as personality. Therefore, non-responder analyses seem to be of use.

There is also a lack of knowledge regarding which particular assessment processes are deemed reasonable for a client. Furthermore, this question of reasonableness also concerns specific genres of psychological instruments, for instance, projective methods, adaptive tests, (easy to see through) personality questionnaires, and objective personality tests.

Researchers are encouraged to engage new (psycho-) technological advances in order to solve traditional problems of psychological assessment; such as merging speed and power, guessing effects at multiple-choice items, public dissemination of item contents, and faking-good or -bad behavior of test-takers.

Psychometricians might deal with improving algorithms for automatic item generation. And they could deal with the problem of a significant test-profile by taking the correlation coefficients of the test characters into account. Of interest would also be which realization of some goodness-of-fit indexes corresponds with the result of a significance test regarding an (IRT-) model's validness – in case any exists, due to specific objectivity *sensu Georg Rasch*.

Statisticians might intensify elaborations of sequential testing approaches; that is for the Welch-test, for methods which have to be applied in case of ordinal-scaled data, for multi-dimensional tests, above all, for discriminant analysis, and e.g. for testing the Rasch-model.

Methodical standards

There are still some improper traditions within psychology concerning statistical analyses. That is, above all a) the “practice of asterisks”, which always implies the highest α of all α -levels one would ever accept – if a researcher tries to impose the result's conclusiveness on the reader by this means, then matter-of-factly quoting the estimated effect size is only informative. b) Furthermore, the arbitrary choice of the type-I-risk (α) without reflecting the consequences of a possible type-II-error is to generally be firmly rejected – instead, calculating *a-posterior* the “result-based type-II-risk” at least approach-

es the state-of-the-art, which is to calculate the sample size: hereby only relevant effects will result in significance, but such relevant effects will not be detected with the probability of some set type-II-risk only (it is most favorable to use the R-routine `seqtest`; Yanagida, 2016). Of course, applying sequential testing is even more preferable; thereby data are sampled one after the other, which generally saves a large part of the sample size, while the type-I- and type-II-risk are fixed (the R-routine `seqtest` serves for such analyses as well). c) Obviously, a significant correlation coefficient is hardly of any use as even a correlation coefficient of .01 can reach significance, given the sample size is large enough – instead, only the determination coefficient is of any meaning (i.e. the effect size in question), with it consequently being better to calculate the sample size in advance based on a certain type-I- and type-II-risk and a value of the determination coefficient that is of practical relevance (at least, the null-hypothesis $H_0: 0 < \rho \leq \rho_0$, e.g. $\rho_0 = .70$, rather than the null-hypothesis $H_0: \rho = 0$ should be tested). Concerning the latter, Schneider, Rasch, Kubinger, and Yanagida (2015) established a sequential (triangular) test of a correlation coefficient's null-hypothesis $H_0: 0 < \rho \leq \rho_0$. The respective computer program for its application is integrated in the R-routine `seqtest` as well.

Finally, researchers are rather encouraged to design their research as an experiment than a survey, because only the former allows causal conclusions.

Postscript: Authors are also warmly encouraged to publish new computer routines (particularly done in R), which support psychological test and assessment modeling.

References

- Brandt, S., Moulton, M., & Duckor, B. (2015). Advances in Rasch modeling: New applications and directions. *Psychological Test and Assessment Modeling*, *57*, 338-341.
- Eckes, T. (2017). Rater effects: Advances in item response modeling of human ratings – Part I. *Psychological Test and Assessment Modeling*, *59*, 443-452.
- Eckes, T. (2018). Rater effects: Advances in item response modeling of human ratings – Part II. *Psychological Test and Assessment Modeling*, *60*, 29-32.
- Etzler, S.L., Rohrmann, S., & Brandt, H. (2014). Validation of the STAXI-2: A study with prison inmates. *Psychological Test and Assessment Modeling*, *56*, 178-194.
- Frey, A., & Kröhne, U. (2012). Guest Editorial: Current issues in educational and psychological measurement: design, calibration, and adaptive testing – Part I. *Psychological Test and Assessment Modeling*, *54*, 363-365.
- Frey, A., König, C., & Spoden, C. (2018a). Advances in Educational Measurement – Part I. *Psychological Test and Assessment Modeling*, *60*, 141-144.
- Frey, A., König, C., & Spoden, C. (2018b). Advances in Educational Measurement – Part II. *Psychological Test and Assessment Modeling*, *60*, 325-326.
- Hothorn, T. & Leisch, F. (2011). Case studies in reproducibility. *Briefings in Bioinformatics*, *12*, 288-300.

- Kastner-Koller, U., & Deimann, P. (2011). Guest Editorial: Caregivers' evaluations of young children's development and behavior. *Psychological Test and Assessment Modeling*, *53*, 212-213.
- Kröhne, U., & Frey, A. (2013). Guest Editorial: Current issues in educational and psychological measurement: design, calibration, and adaptive testing – Part II. *Psychological Test and Assessment Modeling*, *55*, 79-80.
- Kubinger, K.D. (2008). On the revival of the Rasch model-based LLTM: From constructing tests using item generating rules to measuring item administration effects. *Psychology Science Quarterly*, *50*, 311-327.
- Kubinger, K.D., Heuberger, N. & Poinstingl, H. (2010). On the self-evaluation of a journal's impact factor. *Test and Assessment Modeling in Psychology*, *52*, 142-147.
- Pohl, S., & Almann, C. (2015). Missing values in large-scale assessment studies. *Psychological Test and Assessment Modeling*, *57*, 469-471.
- Proyer, R. T., & Ruch, W. (2010). Editorial: Dispositions towards ridicule and being laughed at: Current research on gelotophobia, gelotophilia, and katagelasticism. *Psychological Test and Assessment Modeling*, *52*, 49-59.
- Rasch, D. & Guiard, V. (2004). The Robustness of Parametric Statistical Methods. *Psychology Science*, *46*, 175-208.
- Reeve, B. B., & Teresi, J. A. (2016). Overview to the two-part series: Measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) short forms. *Psychological Test and Assessment Modeling*, *58*, 31-35.
- Rupp, A. A. (2013). A systematic review of the methodology for person fit research in Item Response Theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling*, *55*, 3-38.
- Schneider, B., Rasch, D, Kubinger, K.D., & Yanagida, T. (2015). A Sequential Triangular Test of a Correlation Coefficient's Null-Hypothesis: $0 < \rho \leq \rho_0$. *Statistical Papers*, *56*, 689-699.
- Schorer, J., Baker, J., Stoeger, H., & Ziegler, A. (2015). Identifying effective learning environments – Part I. *Psychological Test and Assessment Modeling*, *57*, 37-39.
- Stadler, M., Greiff, S., & Krolak-Schwerdt, S. (2016). Current methodological issues in educational large-scale assessments – Part I. *Psychological Test and Assessment Modeling*, *58*, 593-595.
- Stadler, M., Greiff, S., & Krolak-Schwerdt, S. (2016). Current methodological issues in educational large-scale assessments – Part II. *Psychological Test and Assessment Modeling*, *59*, 31-33.
- Stemmler, M., & von Eye, A. (2012). Configural Frequency Analysis (CFA) and other non-parametrical statistical methods. *Psychological Test and Assessment Modeling*, *54*, 149-150.
- Stoeger, H., Schorer, J., Baker, J., & Ziegler, A. (2015). Identifying effective learning environments – Part II. *Psychological Test and Assessment Modeling*, *57*, 162.

- Stoeger, H., & Ziegler, A. (2011). Guest Editorial to the special issue: Quantitative approaches to the study of self-regulated learning. *Psychological Test and Assessment Modeling*, 53, 49-50.
- Vialle, W. (2013a). Guest editorial: Current perspectives on the assessment of giftedness – Part I. *Psychological Test and Assessment Modeling*, 55, 247-249.
- Vialle, W. (2013b). Guest editorial: Current perspectives on the assessment of giftedness – Part II. *Psychological Test and Assessment Modeling*, 55, 383-384.
- von Davier, M. (2011). Editorial: Methodological advances in psychological and educational testing. *Psychological Test and Assessment Modeling*, 53, 312-314.
- Yanagida, T. (2016). seqtest: Sequential Triangular Test. R package version 0.1-0. <https://CRAN.R-project.org/package=seqtest> .
- Ziegler, A., Alghawi, M., & Reutlinger, M. (2019, in prep.). New trends in gifted identification. *Psychological Test and Assessment Modeling*, 61.