

Editorial: **Psychological Test and Assessment Modeling**

*Klaus D. Kubinger*¹
(*editor-in-chief*)

This journal has been founded as *Psychologische Beiträge* – a journal for a quarterly review of psychological science research. The scope was broad as in other journals from the 1950s. Therefore, being internationalized in 2003, it became renamed as *Psychology Science*. After the realization that some papers had been cited in the abbreviated manner of *PsycholSci* – so the paper scored on the ISI-indexed well-known journal *Psychological Science*, but not on *Psychology Science* –, such mistakes were tried to prevent renaming the journal again to *Psychology Science Quarterly*, in 2008. However, in the last two years the journal has altered its scope due to the topics of submitted papers and due to (invited) special issues. As a matter of fact, the scope has changed towards the primary subjects of psychology-specific statistical methods and problems, general psychometrics, and psychological assessment in theory and practice.

First of all, the special issue 2008/2 “High Ability Assessment” (cf. Stöger & Ziegler, 2008) should be mentioned. In this issue for instance, Holocher-Ertl, Kubinger, and Hohensinn (2008) dealt with the gap between practical demands and scientific supply when identifying children who may be cognitively gifted. Holling and Kuhn (2008) investigated the factor structure of divergent thinking in gifted children. Phillipson (2008) calibrated underachievers’ achievements using the Rasch model. And Ziegler and Stöger (2008) suggested some content model of learning. There was then a special issue in 2008/3 “The Use of LLTM (Linear Logistic Test Model): Cognitive Modelling and Item-Technology Analysis”. In this issue for instance, Kubinger (2008) brought up a revival of the Rasch model-based LLTM. Embretson and Daniel (2008) used the LLTM for analysing the complexity in mathematical problem solving items; Sonnleitner (2008) and Holling et al. (2008) applied the LLTM for the evaluation of item generating rules (the former for reading comprehension, the latter for statistical wording); Xie and Wilson (2008) used the LLTM for DIF-analyses, and Hohensinn et al. (2008) used this model for

¹ *Correspondence concerning this article should be addressed to:* Prof. Klaus D. Kubinger, PhD, Head of the Division for Psychological Assessment and Applied Psychometrics, University of Vienna, Faculty of Psychology, Liebiggasse 5, 1010 Vienna, Austria; email: klaus.kubinger@univie.ac.at

examining item-position effects. However, there are many other papers to give representative examples of the actual scope of the journal. Whitman et al. (2008) designed an experiment to investigate the fakability of a certain measure of emotional intelligence; Malda et al. (2008) dealt with the adaptation of a cognitive test for a different culture; Teresi et al. (2008) described pertinent methods of disclosing DIF; Koch et al. (2009) evaluated the Critical Incident Technique; Ziegler, Toomela, and Bühner (2009) analysed spurious measurement errors within the Big Five; Platt, Proyer, and Ruch (2009) dealt with the assessment of gelotophobia; Scheiblechner (2009) contrasted Rasch models and so-called pseudo-Rasch models; Dittrich and Hatzinger (2009) illustrated how to fit log-linear Bradley-Terry models for paired comparison; Fill Giordano et al. (2009) attempted to assess learning achievement; Staugaard (2009) evaluated different versions of attentional bias tasks; Foster and Miller (2009) introduced a new format for multiple-choice testing, and finally Kubinger, Rasch, and Yanagida (2009) gave a new suggestion of designing data-sampling for Rasch model calibrating a test. Last but not least, there was a special topic in 2009/4: “Working memory and intelligence” (cf. Schweizer, 2009). There for instance, Unsworth, Spillers, and Brewer (2009) examined some content model of working memory capacity, attention control, and fluid intelligence; Rockstroh and Schweizer (2009) investigated the effect of retest practice.

Hence, we decided to change the title of the journal again, to: *Psychological Test and Assessment Modeling*. There is the possibility that this specialisation leads to a much higher receptiveness and acceptance of the journal – as a matter of fact, we self-evaluated the impact factor of *Psychology Science Quarterly* 2009 according to Kubinger, Heuberger, and Poinstingl (2010, in press), which will be published ongoing. And, of course, some new board members have been acquired, while a few board members have left the board for content reasons.

Nevertheless, the scope of the journal is wide; it publishes important research results on the given topics, including personality psychology. Empirical contributions are welcome, as well as theoretical papers concerning special content models and psychometric or statistical developments. Furthermore, simulation studies to psychometric or statistical models are within the focus of the journal as long as they serve for the solution of basical psychological research questions.

High standards in the application of empirical methods – including elaborated methodological approaches – are desired. Either theory-based experimental designs or the use of excellent methods, procedures, and algorithms are expected. Though some misuse of statistics may be common, this and other improper approaches should be avoided. In particular, one should bear in mind that rejection of the null-hypothesis that a certain correlation coefficient is null, is hardly of any scientific gain, but one should rather use a specific null-hypothesis (cf. Kubinger, Rasch, & Šimečková, 2007). Of course, papers with well-planned studies and experiments that do not deliver significant results, that is to say do not confirm the assumed hypothesis, are in no way excluded; their contribution to scientific knowledge may be just as important as that of papers with positive results.

Besides conventional APA requirements we give the following general hints for authors:

- Use only 5 keywords
- Use “Black and White (2003)” within the text, but “Black & White (2002)” within footnotes, Tables, Figures, or within brackets; please list every author’s name the first time you refer to them, but afterwards, if there are more than three authors, then use Miller et al.
- Don’t use too many abbreviations
- Always use statistical symbols in Italics
- Do not use different type-I-risk levels in your paper for similar questions, but decide in advance for a certain level; there is no need for asterisks in order to indicate the effect size, rather estimate the (relative) effect size
- Do not use alpha-risk, alpha-error, beta-risk, and beta-error, but type-I- and type-II-risk, and -error.
- Do not use N for sample sizes, but n
- Don’t use $T = 21.6$ but $t = 21.6$, because that concerns the t -test and the t -distribution; that is do not follow SPSS
- Please always set a “blank” between statistical expressions and mathematical symbols – that is $p = .03$ instead of $p=.03$
- Don’t use Chi2 or the like, but χ^2
- Please use 50 %, but not 50%
- Don’t use the qualitative categories by Cohen (1960) for the evaluation of the estimated effect size but give the estimated (relative) effect size in its quantitative value
- There is no need to write $p < .001$ and the like, in case SPSS lists $p = .000$ – this is so, because every reader acquainted with mathematics knows that p is not actually zero but the given value is only the result of rounding

We will continue with the tradition of publishing special topics. These, however, will no longer constitute a special issue on their own, but will rather be accompanied by a few *varia* papers. Hence, every issue offers *varia* papers; this is because we can publish a paper very quickly, though there might be a pool of special topics for the forthcoming papers. In the first issue 2010/1 we are glad to publish Part I of the special topic: “Gelotophobia”, edited by the guest editors Proyer and Ruch (2010); Part II will be published in issue 2010/2. There are two special topics in preparation for 2011, “Advances in psychological and educational testing” and “Methods of cluster- and type analyses”. And for 2012 there is a special topic in preparation: “World-wide diversity of intelligence testing”. Of course, any suggestions to the editorial board for other special topics are welcome.

References

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Dittrich, R., & Hatzinger, R. (2009). Fitting loglinear Bradley-Terry models (LLBT) for paired comparisons using the R package *prefmod*. *Psychology Science Quarterly*, 51, 216-242.
- Embretson, S. E., & Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science Quarterly*, 50, 328-344.
- Fill Giordano, R., Litzenberger, M., Wagner-Menghin, M., & Binder, J. (2009). Assessing learning achievement, time effort, learning approaches and tempo during learning within the experiment-based behavioral task LAsO – reliability and incremental validity. *Psychology Science Quarterly*, 51, 247-265.
- Foster, D., & Miller, H. L. (2009). A new format for multiple-choice testing: Discrete-Option Multiple-Choice. Results from early studies. *Psychology Science Quarterly*, 51, 355-369.
- Hohensinn, C., Kubinger, K. D., Reif, M., Holocher-Ertl, S., Khorramdel, L., & Frebort, M. (2008). Examining item-position effects in large-scale assessment using the Linear Logistic Test Model. *Psychology Science Quarterly*, 50, 391-402.
- Holling, H., Blank, H., Kuchenbäcker, K., & Kuhn, J.-T. (2008). Rule-based item design of statistical word problems: A review and first implementation. *Psychology Science Quarterly*, 50, 363-378.
- Holling, H., & Kuhn, J.-T. (2008). Does intellectual giftedness affect the factor structure of divergent thinking? Evidence from a MG-MACS analysis. *Psychology Science Quarterly*, 50, 283-294.
- Holocher-Ertl, S., Kubinger, K. D., & Hohensinn, C. (2008). Identifying children who may be cognitively gifted: the gap between practical demands and scientific supply. *Psychology Science Quarterly*, 50, 97-111.
- Koch, A., Strobel, A., Kici, G., & Westhoff, K. (2009). Quality of the Critical Incident Technique in practice: Interrater reliability and users' acceptance under real conditions. *Psychology Science Quarterly*, 51, 3-15
- 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, in print). On the self-evaluation of a journal's impact factor. *Test and Assessment Modeling in Psychology*, 52.
- Kubinger, K. D., Rasch, D., & Šimečková, M. (2007). Testing a correlation coefficient's significance: Using $H_0: 0 < \rho \leq \lambda$ is preferable to $H_0: \rho = 0$. *Psychology Science*, 49, 74-87.
- Kubinger, K. D., Rasch, D., & Yanagida, T. (2009). On designing data-sampling for Rasch model calibrating an achievement test. *Psychology Science Quarterly*, 51, 370-384.

- Malda, M., van de Vijver, F. J. R., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly, 50*, 451-468.
- Phillipson, S. (2008). The optimal achievement model and underachievement in Hong Kong: an application of the Rasch model. *Psychology Science Quarterly, 50*, 147-172.
- Platt, T., Proyer, R., & Ruch, W. (2009). Gelotophobia and bullying: The assessment of the fear of being laughed at and its application among bullying victims. *Psychology Science Quarterly, 51*, 135-147.
- Rockstroh, S., & Schweizer, K. (2009). An investigation of the effect of retest practice on the relationship between speed and ability in attention, memory and working memory tasks. *Psychology Science Quarterly, 51*, 420-431.
- Scheiblechner, H. H. (2009). Rasch and pseudo-Rasch models: suitability for practical test applications. *Psychology Science Quarterly, 51*, 181-194.
- Schweizer, K. (2009). Editorial for special topic: Working memory and intelligence. *Psychology Science Quarterly, 51*, 385-387.
- Sonnleitner, P. (2008). Using the LLTM to evaluate an item-generating system for reading comprehension. *Psychology Science Quarterly, 50*, 345-362.
- Staugaard, S. R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science Quarterly, 51*, 339-350.
- Stöger, H., & Ziegler, A. (2008). Editorial: High Ability Assessment. *Psychology Science Quarterly, 50*, 91-96.
- Teresi, J. A., Ramirez, M., Lai, J.-S., & Silver, S. (2008). Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. *Psychology Science Quarterly, 50*, 538-612.
- Unsworth, N., Spillers, G. J., & Brewer, G. A. (2009). Examining the relations among working memory capacity, attention control, and fluid intelligence from a dual-component framework. *Psychology Science Quarterly, 51*, 388-402.
- Whitman, D. S., van Rooy, D. L., Viswesvaran, C., & Alonso, A. (2008). The susceptibility of a mixed model measure of emotional intelligence to faking: a Solomon four-group design. *Psychology Science Quarterly, 50*, 44-63.
- Xie, Y., & Wilson, M. (2008). Investigating DIF and extensions using an LLTM approach and also an individual differences approach: an international testing context. *Psychology Science Quarterly, 50*, 403-416.
- Ziegler, A., & Stöger, H. (2008). A learning oriented subjective action space as an indicator of giftedness. *Psychology Science Quarterly, 50*, 222-236.
- Ziegler, A., Toomela, A., & Bühner, M. (2009). A reanalysis of Toomela (2003): Spurious measurement error as cause for common variance between personality factors. *Psychology Science Quarterly, 51*, 65-75.