The Journal of Open Source Software

measr: Bayesian psychometric measurement using Stan

W. Jake Thompson [1,2]

1 Accessible, Teaching, Learning, and Assessment Systems **2** University of Kansas

DOI: 10.21105/joss.05742

Software

- Review
- Repository
- Archive

Editor: Chris Hartgerink

Reviewers:

- @smasongarrison
- @Wenchao-Ma
- @seandamiandevine

Submitted: 14 July 2023
Published: 31 October 2023

Summary

In educational and psychological research, we are often interested in discrete latent states of individuals responding to an assessment (e.g., proficiency or non-proficiency on educational standards, the presence or absence of a psychological disorder). Diagnostic classification models (DCMs; also called cognitive diagnostic models [CDMs]) are a type of psychometric model that facilitates these inferences (Rupp et al., 2010; von Davier & Lee, 2019). DCMs are multi-dimensional, meaning that we can classify respondents on multiple latent attributes within a profile of skills. A Q-matrix is then used to define which items on the assessment measure each attribute. Using the pre-defined latent profiles and the Q-matrix, DCMs then estimate the probability that respondents are in profile, or have the corresponding pattern of proficiency, or presence, of the attributes. This means that DCMs are able to provide fine-grained feedback on specific skills that may need additional instruction in an educational context, or particular symptoms that may be contributing to a diagnosis in a psychological context. Finally, because DCMs are classifying respondents rather than placing them along a performance continuum, these models are able to achieve more reliable results with shorter test lengths (Templin & Bradshaw, 2013), reducing the burden on respondents.

Given these benefits, the goal of measr is to make DCMs more accessible to applied researchers and practitioners by providing a simple interface for estimating and evaluating DCMs.

Statement of need

measr is an R package developed to easily estimate and evaluate DCMs in applied settings. Despite the ability of DCMs to provide reliable, fine-grained feedback on specific skills, these models have not been widely used for research or operational programs. This is due in large part to limitations in existing software for estimating and evaluating DCMs (Ravand & Baghaei, 2020; Sessoms & Henson, 2018). Typically, DCMs are estimated with a maximum likelihood estimator and then evaluated using limited-information fit indices (e.g., Liu et al., 2016). This is the approach taken when using Mplus (e.g., Templin & Hoffman, 2013) and popular R packages GDINA (Ma & de la Torre, 2020) and CDM (George et al., 2016). However, as the name "limited-information" implies, these methods only look at limited relationships between the items, such as univariate or bivariate relationships. This means that higher-level relationships between the items cannot be evaluated (e.g., relationships between triplets of items).

Bayesian estimation methods offer more robust methods for evaluating model fit through posterior predictive checks (Park et al., 2015; Thompson, 2019). To date, there are three R packages that offer Bayesian estimation of DCMs: dina (Culpepper, 2015), hmcdm (Zhang et al., 2023), and blatent (Templin, 2020). However, all of these packages only estimate a single type of DCM, severely limiting their generalizability to a wide range of applications.

The measr package seeks to overcome the limitations of existing software options by serving as an interface to the Stan probabilistic programming language (Carpenter et al., 2017). With Stan as a backend, measr can estimate a wide variety of DCMs. Primarily, measr supports the estimation of the loglinear cognitive diagnostic model (LCDM). However, because the LCDM is a general DCM that subsumes many subtypes (Henson et al., 2008), measr also supports other DCMs such as the deterministic inputs, noisy "and" gate (DINA) model (de la Torre & Douglas, 2004) and the deterministic inputs, noisy "or" gate (DINO) model (Templin & Henson, 2006). After estimation, measr provides model evaluations using both limited-information indices and posterior predictive checks. By providing straightforward estimation and evaluation of DCMs, measr makes these models more accessible to practitioners and applied researchers. Thus, with measr, users get the power of Bayesian methods for model evaluation, compatibility with other packages in the larger Stan ecosystem, and a user-friendly interface so that knowledge of the Stan language is not required. However, models estimated with measr also include the fitted Stan object, so users can access it if they are familiar with Stan and prefer to work with that object. Additionally, the Stan code used to estimate the model is also returned so that users familiar with the Stan language can use that code as a starting point for writing their own customized models.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D210045 to the University of Kansas. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

We are grateful to the project advisory committee members who provided feedback on the development of the R package: Russell Almond, Claudia Flowers, Robert Henson, and Matthew Madison.

References

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics*, *40*(5), 454–476. https://doi.org/10.3102/1076998615595403

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333–353. https://doi.org/10.1007/BF02295640

George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, *74*(2), 1–24. https://doi.org/10.18637/jss.v074.i02

Henson, R. A., Templin, J., & Willse, J. T. (2008). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191–210. https://doi.org/10.1007/s11336-008-9089-5

Liu, Y., Tian, W., & Xin, T. (2016). An application of M_2 statistic to evaluate the fit of cognitive diagnostic models. *Journal of Educational and Behavioral Statistics*, *41*(1), 3–26. https://doi.org/10.3102/1076998615621293

Ma, W., & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software*, *93*(14), 1–26. https://doi.org/10.18637/jss.v093.i14

Park, J. Y., Johnson, M. S., & Lee, Y.-S. (2015). Posterior predictive model checks for cognitive diagnostic models. *International Journal of Quantitative Research in Education*, *2*(3–4), 244–264. https://doi.org/10.1504/IJQRE.2015.071738

Ravand, H., & Baghaei, P. (2020). Diagnostic classification models: Recent developments, practical issues, and prospects. *International Journal of Testing*, *20*(1), 24–56. https://doi.org/10.1080/15305058.2019.1588278

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press. ISBN: 978-1-60623-527-0

Sessoms, J., & Henson, R. A. (2018). Applications of diagnostic classification models: A literature review and critical commentary. *Measurement: Interdisciplinary Research and Perspectives*, *16*(1), 1–17. https://doi.org/10.1080/15366367.2018.1435104

Templin, J. (2020). *blatent: Bayesian latent variable models*. https://CRAN.R-project.org/package=blatent

Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, *30*(2), 251–275. https://doi.org/10.1007/s00357-013-9129-4

Templin, J., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

Templin, J., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using mplus. *Educational Measurement: Issues and Practice*, *32*(2), 37–50. https://doi.org/10.1111/emip.12010

Thompson, W. J. (2019). *Bayesian psychometrics for diagnostic assessments: A proof of concept* (Research Report No. 19-01). University of Kansas; Accessible Teaching, Learning, and Assessment Systems. https://doi.org/10.35542/osf.io/jzqs8

von Davier, M., & Lee, Y.-S. (Eds.). (2019). *Handbook of diagnostic classification models: Models and model extensions, applications, software packages*. Springer Cham. https://doi.org/10.1007/978-3-030-05584-4

Zhang, S., Wang, S., Chen, Y., & Kwon, S. (2023). *hmcdm: Hidden Markov cognitive diagnosis models for learning*. https://CRAN.R-project.org/package=hmcdm