Getting Started with the Graded Response Model (GRM): A gentle introduction and tutorial in R

**Abstract**

This tutorial paper will provide an introduction to the Graded Response Model (GRM), a tool for testing measurement validity under the Item Response Theory (IRT) paradigm. Addressing common problems of measurement imprecision and lack of construct validity, the tutorial guides researchers through a one-dimensional GRM analysis in the R environment, using mirt and ggmirt packages. GRM is specifically designed to examine the psychometric properties of psychological scales with polytomous items. The tutorial will illustrate the procedure using data from the Open Psychometrics Database on the Right-Wing Authoritarianism (RWA) scale, outlining the theoretical underpinnings of GRM, and steps for data preparation, model fitting, interpretation of results, and dealing with common issues and anomalies that may typically arise in the process.

Keywords: graded response model, item response theory, mirt, ggmirt, R.

**Getting Started with the Graded Response Model (GRM): A gentle introduction and tutorial in R**

Measurement plays a fundamental role in psychological research. Nevertheless, imprecise measurement and lack of evidence for construct validity are widespread (Barry et al., 2014; Borsboom, 2006; Weidman et al., 2017), which in turn, can undermine validity in all aspects of research (Flake & Fried, 2020). Construct validity is often routinely tested with unrealistic assumptions, such as that the latent trait being measured, and its corresponding measurement error remain the same for all individuals regardless of their position on the latent trait spectrum (i.e., “classical test theory” or CTT). In this tutorial, I will attempt to gently introduce a Graded Response Model (GRM) analysis to the readers that allows researchers to test their measurement validity with assumptions untested in the routine CTT procedures.

This tutorial paper will seek to guide applied researchers through the process of conducting a one-dimensional GRM analysis in the R environment, specifically using the *mirt* (Chalmers, 2012) and *ggmirt* (Masur, 2022/2023) packages. GRM is a part of the IRT family, which is specifically designed to test psychological scales with polytomous (i.e., ordinal, Likert-style) items. As a popular analytical tool in psychometrics, GRM serves as an effective modeling strategy for estimating the likelihood of an individual’s response to polytomous items, thereby providing deeper insights into the measured psychological constructs.

The primary goal of this tutorial is to introduce the application of GRM in a real-world scenario and to highlight the practical benefits of this approach. To this end, I will demonstrate the process of examining the psychometric properties of the Right Wing Authoritarianism (RWA) scale (Altemeyer, 1981) using the data the from Open Psychometrics Database. As the RWA scale is a popular measurement tool commonly used in social psychological research, this paper will demonstrate the advantage and relevance of GRM analysis in examining psychometric properties of a personality scale.

This tutorial will begin by outlining the theoretical underpinnings of GRM and emphasizing its unique ability to handle graded responses that are often used in psychological measurement. This paper will delineate the differences between GRM and other IRT models, then highlight situations where GRM is most advantageous. We will then move into a hands-on, step-by-step guide to performing GRM analyses using the *mirt* and *ggmirt* packages in R.

After introducing the R packages, the tutorial will walk researchers through the essential steps of importing data, preparing it for analysis, testing the unidimensionality assumption by inspecting local dependence statistics, fitting the graded response model, evaluating the model fit, and interpreting the results. Special emphasis will be placed on interpreting the model parameters of the GRM and understanding their implications for research findings. The tutorial will then delve into how to handle and interpret common issues and anomalies that may arise during the analysis process.

Finally, the tutorial will conclude with a comprehensive discussion of the findings from the RWA scale, integrating the technical, theoretical, and practical insights gleaned from the GRM analysis. By synthesizing these insights, researchers will be better equipped to apply the GRM in their future studies.

In conclusion, this tutorial paper will provide a practical, detailed, and comprehensive guide for applied researchers to implement a GRM analysis using *mirt* and *ggmirt* packages in R. By using a real-world dataset as a case study, this tutorial will aim to keep the understanding and application of GRM grounded in practice, ultimately helping applied researchers to improve the quality of their research.

**REFERENCES**

Altemeyer, B. (1981). *Right-wing authoritarianism*. The University of Manitoba Press.

Barry, A. E., Chaney, B., Piazza-Gardner, A. K., & Chavarria, E. A. (2014). Validity and Reliability Reporting Practices in the Field of Health Education and Behavior: A Review of Seven Journals. *Health Education & Behavior*, *41*(1), 12–18. https://doi.org/10.1177/1090198113483139

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425–440. https://doi.org/10.1007/s11336-006-1447-6

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Masur, P. K. (2023). *ggmirt* [R]. https://github.com/masurp/ggmirt (Original work published 2022)

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, *17*(2), 267–295. https://doi.org/10.1037/emo0000226