Book Review of *Rasch Models for Solving Measurement Problems: Invariant Measurement in the Social Sciences* by Engelhard Jr. and Wang

As a paradigm shift from classical test theory, Rasch measurement theory introduces a framework of scales that is capable of producing invariant measurement, which plays an increasingly important role in solving practical measurement problems in the social sciences, psychometrics, and health sciences. Researchers have increasingly adopted Rasch measurement to develop and validate scales in a variety of educational and psychological contexts (e.g. Mendoza & Yan, 2021; Yan & Pastore, 2022). Rasch measurement is often used in the validation of scales in conjunction with structural equation models (SEM), which provide meaningful answers to a variety of measurement questions. However, it remains unfamiliar to many researchers, practitioners, and educators who have embraced classical test theory.

Rasch measurement theorists support invariance philosophically and theoretically through the metaphor of a ruler, including invariance across persons and invariance across items. It ensures that persons with greater ability will do better regardless of which item is selected and that the chances of success on easy items must be greater than on difficult items, regardless of the person's ability. The book under review provides accessible answers to the methodological steps involved in scale construction, evaluation, maintenance, and use, building upon Rasch measurement theory. For illustrating ideas and demonstrating techniques, the Food Insecurity Experience (FIE) scale is used, which is composed of eight dichotomous items. On the book's website, the authors provide the Facets syntax and R code for the ERMA program to evaluate the model-data fit and obtain parameter estimates.

Chapter 1 discusses the theoretical underpinnings of Rasch invariant measurement. Rasch's (1961) recognition that measurement should be focused on individuals rather than groups, as well as his specific requirements for achieving objectivity in individual-centered measurement, is a major contribution to the field of measurement. Also discussed in the chapter is the difference between Rasch measurement theory and the IRT model, which emphasizes that Rasch measurement theory prioritizes meeting the model's requirements and is confirmatory in nature, requiring data that fits the model, in contrast to the IRT model, which focuses on reproducing a specific dataset and attempting to find a model that fits the data. Additionally, the chapter discusses four measurement issues: the selection and definition of latent variables, differential item functioning, interchangeability of items, and standard setting. By applying Rasch measurement theory to these issues, it is possible to create invariant measures, person-independent item calibrations, as well as meaningful and comparable scales.

Chapter 2 examines how the requirements of invariant measurement can be used to guide the construction of scale parameters at every level. Four building blocks are necessary for scale construction based on Rasch measurement theory: specifying the latent variables to be measured, designing an observational design (items or questions), specifying scoring rules, and applying the Rasch model to the observational data to create an empirical Wright Map. The authors provide additional technical information on utilizing Rasch models to estimate item and person locations, along with resources for analyzing data sets.

Chapter 3 discusses the details of assessing scales, including item-invariant person measures and person-invariant item calibration, emphasizing that invariant measures can only be achieved when a good model-data fit is obtained. It describes fit indices and procedures for detecting misfit items and personnel for model-data fit, which includes residual statistics and the Infit and Outfit statistics, and discusses the reliability of the separation index. To achieve invariant measurement, the authors emphasize that items must be calibrated invariantly across subgroups of persons. This ensures that the rank ordering of items is invariant across subgroups and that the item difficulty estimates are invariant throughout the range of sampling variation. There is also a discussion of differential item functioning, a problem that can affect the reliability, validity, and fairness of a Rasch scale, as well as the importance of unidimensionality and local independence in item response theory.

Chapter 4 focuses on the comparability of person scores and describes the maintenance of scales for the Rasch model, particularly the content of item interchangeability. The chapter emphasizes the importance of maintaining a stable and invariant representation of a latent variable regardless of the specific items or scales used for measurement. It discusses the concept of comparability and its significance in obtaining comparable scores from different subsets of items or scales. The goal is to establish an invariant metric for the latent variable, ensuring that persons measured by different instruments can be compared.

Chapter 5 introduces three testing foundations for the use of person scores obtained from a Rasch scale in accordance with the Standards for Educational and Psychological Testing (Test Standards; AERA, APA, & NCME, 2014): validity, reliability, and fairness. An emphasis is placed on the significance of invariant measurement and its relationship to these Test Standards. The chapter discusses how scores obtained from the Rasch scales can be used for various purposes, including setting meaningful, ordered categories to guide policy decisions. Rasch theory supports the intended use of

Book Review of Rasch Models for Solving Measurement Problems: Invariant Measurement in the Social Sciences by Engelhard Jr. and Wang

307

test scores through examples of different procedures and cut scores used in the assessment of educational achievement.

Chapter 6 summarizes the main points made by the authors, emphasizing the importance of invariant measurement in creating reliable and valid Rasch scales. Furthermore, the chapter identifies various domains in which Rasch measurement theory can be applied. Additionally, it discusses the family of Rasch models, including Dichotomous, Partial Credit, Rating Scale, Binomial Trials, Poisson Counts, and Facets, as well as extensions such as mixed models, multilevel models, and multidimensional models. With these extensions, Rasch measurement theory can be applied to a variety of complex measurement problems with added flexibility and applicability.

In my capacity as a reviewer, the book caught my attention mainly based on three factors: the reader-friendliness of the content, the standardization of the methodology, and the transparency of the data. It is important to note, in the first place, that the book has been written in a way that addresses the needs of those readers who do not have a specialized background in the field. It provides a brief introduction to philosophical and theoretical foundations, illustrating how Rasch invariant measurement differs from classical test theory. Compared to classic introductory readings on Rasch, such as Bond et al. (2020), Engelhard (2013), and Engelhard and Wind (2017), this book describes relevant concepts and technical details in layman's terms. The majority of its chapters center around the Rasch scale, thereby rendering the content particularly suitable for novice readers. Secondly, this book provides a step-by-step approach to constructing, evaluating, maintaining, and using scales. As a result, readers with similar interests are able to gain a better understanding of the relevant operations and analyses. By elaborating on the FIE scale, this book provides greater depth and detail than some articles that use Rasch to develop and validate scales (e.g., Linacre, 2002). As a third aspect, the authors provide both the Facets syntax and the R code for the ERMA analysis, to facilitate the analysis of data for readers who are not familiar with computer programs, such as Facets and R. Sample data and program syntax in this book will encourage readers to experiment with Rasch scales initially, thereby encouraging the wider application of invariant measurement.

Readers may be misled by this book's title into believing that it is a comprehensive introduction to various Rasch models; however, the authors provide an excellent theoretical introduction and practical guidance on how to measure Rasch scales invariantly that is straightforward to implement. As an essential introduction to the invariant measurement of Rasch scales, this book is highly recommended as a supplementary text for postgraduate courses on measurement and

for researchers who are interested in Rasch but lack relevant knowledge.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge.

Engelhard Jr, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.

Engelhard Jr, G., & Wind, S. (2017). *Invariant measurement with raters and rating scales: Rasch models for rater-mediated assessments*. Routledge.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of applied measurement, 3*(1), 85-106.

Mendoza, N. B., & Yan, Z. (2021). Validation of a subject-specific student self-assessment practice scale (SaPS) among secondary school students in the Philippines. *Journal of Psychoeducational Assessment, 39*(4), 481-493. https://doi.org/10.1177/0734282921994374

Rasch, G. (1961). On general laws and meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 321–333). Berkeley, CA: University of California Press.

Yan, Z., & Pastore, S. (2022). Assessing teachers' strategies in formative assessment: The teacher formative assessment practice scale. *Journal of Psychoeducational Assessment, 40*(5), 592-604. https://doi.org/10.1177/07342829221075121

Xiao Xie*

Universiti Putra Malaysia, Malaysia

Corresponding author: Xiao Xie,
E-mail: gs58879@student.upm.edu.my