

A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research

Abstract

Over the past decades, the application of Rasch measurement in language assessment has gradually increased. In the present study, 215 papers using Rasch measurement published in 21 applied linguistics journals were reviewed and coded for multiple features. It was found that seven Rasch models and 23 software packages were adopted in these papers, with many-facet Rasch measurement ($n=100$) and *Facets* ($n=113$) being the most frequently used Rasch model and software, respectively. Significant differences were detected between the number of papers that applied Rasch measurement to different language skills and components, with writing ($n=63$) and grammar ($n=12$) being the most and least frequently investigated, respectively. In addition, significant differences were found between the number of papers reporting person separation ($n=73$, not reported: $n=142$) and item separation ($n=59$, not reported: $n=156$) and those that did not. An alarming finding was how few papers reported unidimensionality check ($n=57$ vs 158) and local independence ($n=19$ vs 196). Finally, a multilayer network analysis revealed that research involving Rasch measurement has created two major discrete communities of practice (clusters), which can be characterized by features such as language skills, the Rasch models used, and the reporting of item reliability/separation vs person reliability/separation. Cluster 1 was accordingly labelled the production and performance cluster, whereas cluster 2 was labelled the perception and language elements cluster. Guidelines and recommendations for analyzing unidimensionality, local independence, data-to-model fit, and reliability in Rasch model analysis are proposed. Finally, it is argued that the replicability of item and person measures, fit, unidimensionality, local independence, and reliability/separation estimates in studies involving Rasch measurement should be investigated in future research. Rasch measurement allows for reproducibility analysis by offering researchers the opportunity to conduct differential item/facet/distractor/rater functioning. Such analyses would reveal whether measurement tools maintain their properties under different conditions.