

Initial Attempts to Clean Up the Messy Middle Problem

Friday 25 July 2025 08:45 (15 minutes)

Oral presentation

Initial Attempts to Clean Up the Messy Middle Problem

Author

Jimmy de la Torre

Affiliation

University of Hong Kong

Abstract

Accurately measuring learning growth is a crucial exercise because it provides various stakeholders with a comprehensive picture of students' progress and school quality over time. Recently, learning progression (LR) has been introduced as a framework to understand what students know at multiple levels of learning across different time points. Although promising, as LP allows for the integration of learning sciences, modern psychometrics, and rigorous assessment design, many issues therein remain unsolved. Chief among them is the issue of the messy middle (MM) because it directly affects the validity of inferences from LR-based assessments. MM refers to the muddled item difficulties, as well as student abilities at the intermediate levels of the progression. Previous studies have shown that MM, which can be construed as departures from expected item difficulty ordering, and in some instances, item difficulty categories, can result from fitting item response theory (IRT) models that may be too simple for the data. Consequently, in this study, a more complex IRT model, specifically, the four-parameter logistic model (4PL), is explored as a means to address the MM problem. A simulation study involving 15 items, where items 1-5, 6-10, and 11-15 were categorized as easy, medium, and difficult, respectively, was conducted. Items 5, 6, 10, and 11 were designated as boundary items (i.e., items at the boundary of a difficulty category), and the rest as nonboundary items. Moreover, items 5 and 11 (6 and 10) are designated as outer (inner) boundary items. Nonboundary item responses were generated using the 1PL, whereas boundary item responses using the 4PL. The item difficulty parameters were uniformly spaced from -1.4 to 1.4; furthermore, the discrimination parameters (a) of all items, as well as the guessing parameters (g) of items 6 and 11, and slip parameters (s) of items 5 and 10, were manipulated. To examine in an ideal condition how fitting the 1PL to 4PL data produces switches in the boundary item difficulty categorizations, which presupposes switches in item difficulty parameter estimates, the sample size was fixed at 100,000. It was found that the guessing and slip parameters interact with the discrimination parameters to produce category switches. In particular, if the discrimination is low (i.e., $a = 0.5$), category switches occurred when the inner guessing or slip parameter was high (i.e., $g_6, s_{10} \geq .18$); if the discrimination is high (i.e., $a = 2.0$), category switches occurred when the outer slip or guessing parameter was moderately high (i.e., $s_5, g_{11} \geq .10$); however, when the discrimination is average (i.e., $a = 1.0$), category switches occurred only when $s_5, g_{11} \geq .04$ and $g_6, s_{10} \geq .08$. In general, category switches can be represented by four regions of a coordinate system, where the x-axis represents the discrimination parameter, and the y-axis the guessing or slip parameter. Depending on the region, outer or inner items or both can either switch or not switch categories. No switches in the item difficulty estimates, hence, no item difficulty category switches were observed when the 4PL was fitted to the data.

Keywords

Messy middle; Learning progression; 4PL

Primary authors: DE LA TORRE, Jimmy (University of Hong Kong); Ms LUO, Lin (University of Hong Kong); Dr WU, Xiaopeng (University of Hong Kong); Mr FENG, Zechu (University of Hong Kong)

Presenters: DE LA TORRE, Jimmy (University of Hong Kong); Ms LUO, Lin (University of Hong Kong); Dr WU, Xiaopeng (University of Hong Kong); Mr FENG, Zechu (University of Hong Kong)

Session Classification: Session 18 : "Latent factors and errors in psychometric measurement"

Track Classification: Measurement: Measurement