ACTRC Assessment Curriculum & Technology Research Centre

Analyzing Teacher Assessment Competence:
Insights from the SOLO Taxonomy and Item Response Theory

Louie Cagasan
Kevin Carl Santos

August 30, 2024

National Conference on Educational Measurement and Evaluation
The Verdure, 4th Floor, Henry Sy Sr. Hall, DLSU, Manila

The Assessment, Curriculum and Technology Research Centre is a partnership between the University of Melbourne and the University of the Philippines supported by Australian Government.

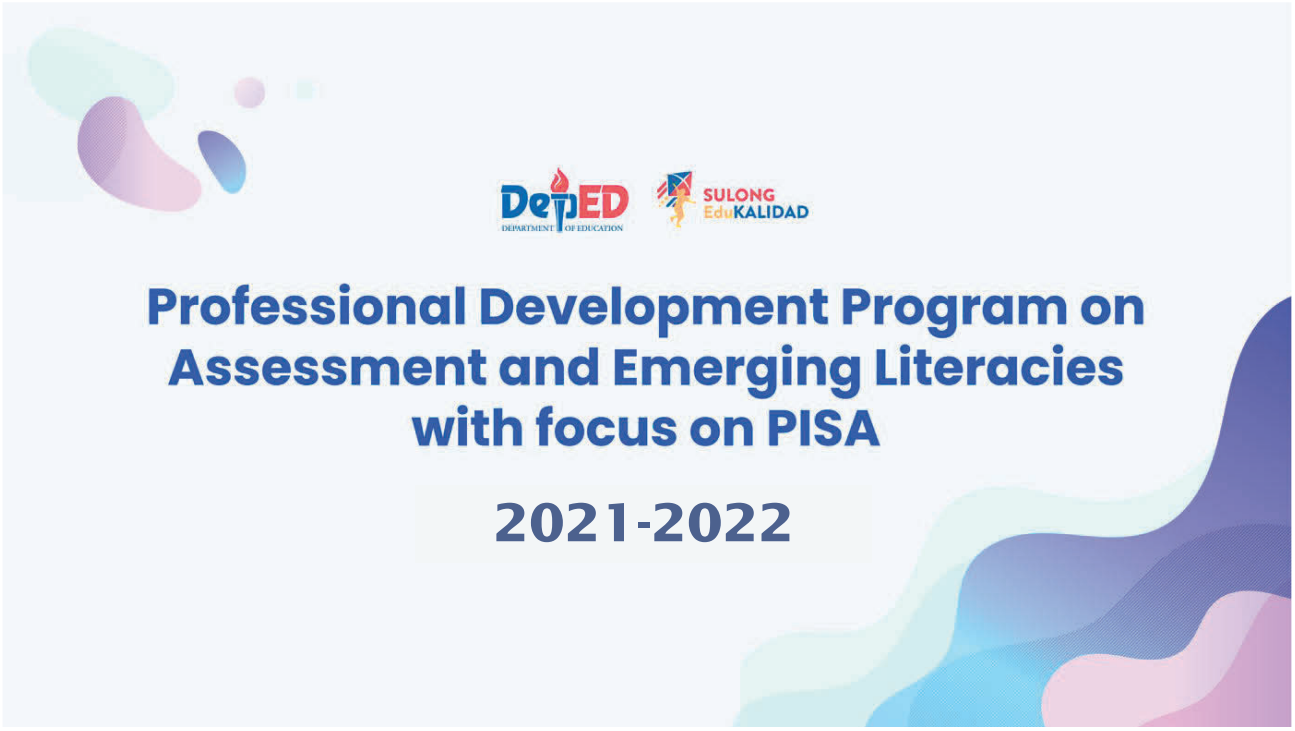
Australian Aid

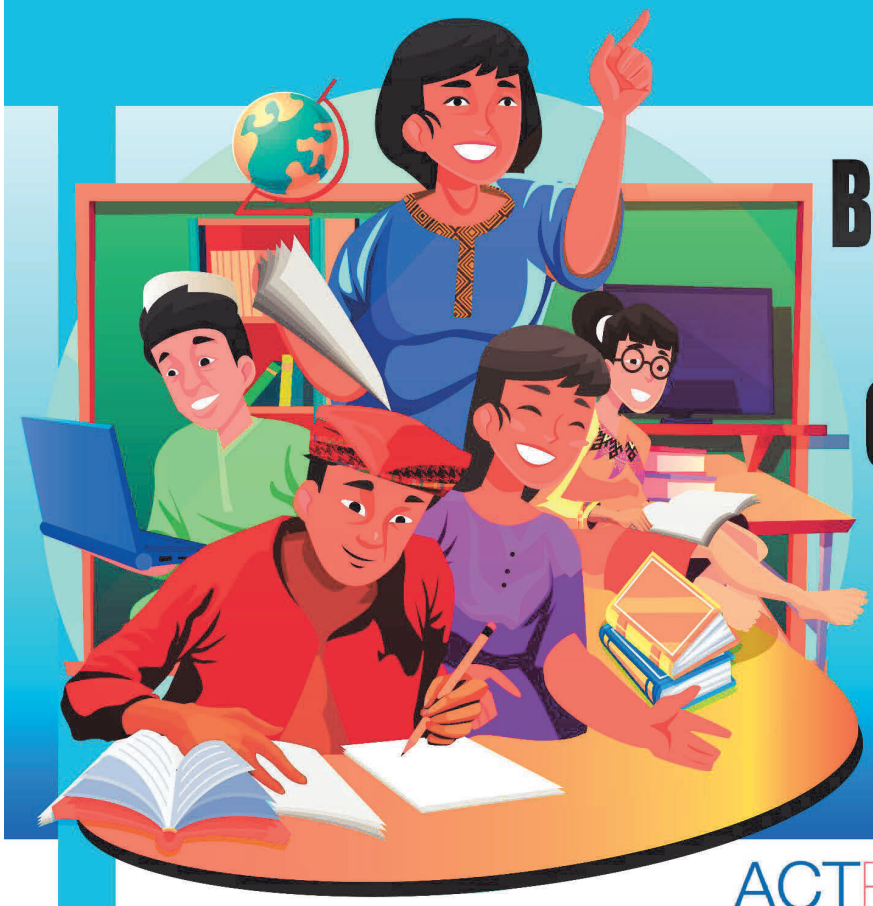
Suggested Citation

Cagasan, L., & Santos, K. (2024, August 29-31). Analyzing Teacher Assessment Competence: Insights from the SOLO Taxonomy and Item Response Theory [Conference panel presentation]. National Conference on Educational Measurement and Evaluation 2024, Manila, Philippines.

Outline

I. Assessment Professional Development Activity
II. Objective of the Presentation
III. Scope and Analysis
IV. Results
V. Discussion





Building Basic Assessment Competence of Teachers

ACTRC Australian Aid THE UNIVERSITY OF MELBOURNE

Assessment Professional Development for Teachers

Building Basic Assessment Competence

CAGASAN, L., BUSTOS, T., FERIDO, M., DE CASTRO, X., NEPOMUCENO, J., BAGUI, L., DELA CRUZ, J., LOBERIZA, K., SANTOS, K., ROBERTSON, P., & RICKARDS, F.

ACTRC

ACTRC is a partnership between the University of Melbourne and the University of the Philippines, supported by the Australian Government.

- The target participants of the program were Math, Science, and English teachers in 9,344 Junior High Schools (JHS) nationwide or 28,032 JHS teachers.
- One of the goals of the program was to improve teachers' assessment literacy and ACTRC delivered the module entitled, **Building Basic Assessment Competence of Teachers** (with six sessions).
- At the end of the module, participants were expected to understand that assessment is a continuous process of collecting, interpreting, and using evidence of students' performance to adjust and improve their teaching.

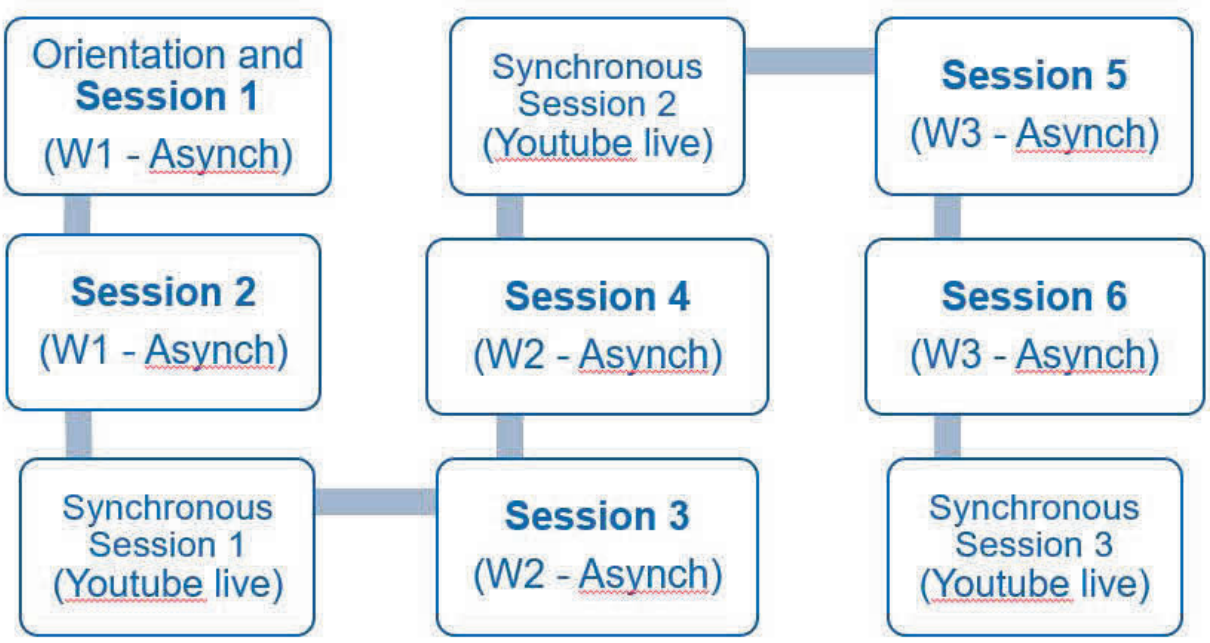
Sessions	Objectives
S1: Assessment and its Purposes	• Recognize the important role of assessment in the teaching-learning process • Determine whether the assessment is formative or summative • Differentiate between formal and informal formative assessment
S2: Assessment Planning	• Explain the importance of planning for assessment • Enumerate the different factors considered when planning assessments • Describe the steps involved in assessment planning
S3: Evidence Collection	• Demonstrate understanding of the importance of a progression of learning in designing assessments • Relate a student's zone of proximal development (ZPD) to assessment design • Create tasks or items assessing students' ZPD with respect to a progression of learning

Sessions	Objectives
S4: Interpretation and Use of Student Data during Classroom Discussion	• Identify where the student is in a progression of learning • Identify the gap between the desired learning outcome and students' current understanding • Provide ways in guiding students to achieve learning outcomes when found to be in the zone of proximal development • Identify their current practice in the four levels of Formative Assessment Practice
S5: Interpretation and Use of Student Data between Lessons	• Interpret and use hypothetical assessment data in "between lessons" or "day by day assessment cycles" • Use assessment interpretation to help learners progress to higher developmental levels • Apply the Three Feedback Questions and Four Levels of Feedback to provide appropriate feedback to learners.
S6: Evaluation of the Assessment Process	• Enumerate considerations in evaluating the assessment process • Apply key principles in evaluating the assessment process • Reflect on assessment practices and identify areas for improvement

Delivery

- The module ran for a total of 18 hours, distributed over three weeks.
- **Asynchronous sessions** involved watching short lecture videos, reading related materials, answering quizzes with built-in feedback, and writing reflection logs. **Evaluation questions were given at the end.**
- **Synchronous sessions** (via YouTube Live) included a summary of the previous sessions and responses to participants' questions during the session.

Delivery



Number of Evaluation Items

APD	Multiple-Choice Items	SOLO Items	Total
Session 1	0	3	3
Session 2	5	0	5
Session 3	5	0	5
Session 4	0	3	3
Session 5	4	3	7
Session 6	3	1	4
Total	**17**	**10**	**27**

Objectives

- To explore the item characteristics of the quiz data using Item Response Theory (IRT) models
- To evaluate the hypothesized order or categories of items based on **Structure of Observed Learning Outcomes** (SOLO) taxonomy (Biggs and Collis,1982)

Scope

- Science teachers
- Only those with complete quiz data (n=1574) were included in the analysis.

Analysis

- IRT Analysis using the MIRT package in R
  - "Mixed" model - 2-Parameter Logistic and Graded Response Models
  - "Rasch" model - Rasch and Partial Credit Models
- Model comparison

Number of Evaluation Items

APD	Multiple-Choice Items	SOLO Items	Total
Session 1	0 (2)	3 (1)	3
Session 2	5	0	5
Session 3	5	0	5
Session 4	0	3	3
Session 5	4	3	7
Session 6	3	1	4
Total	**17 (19)**	**10 (8)**	**27**

Model Fit (Test Level)

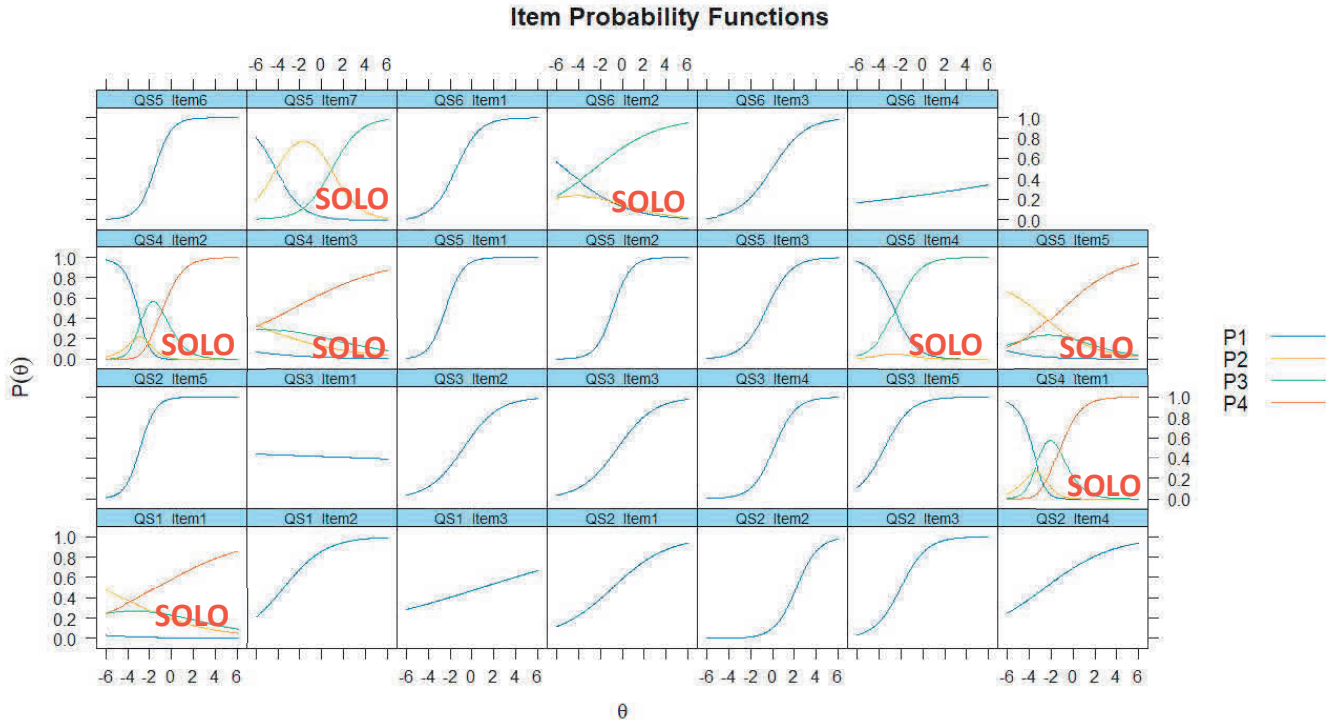
Models	AIC	SABIC	HQ	BIC	logLik
Mixed	51369.89	51516.26	51503.39	51729.11	-25618
Rasch	51810.92	51900.49	51892.61	52030.73	-25864.5

The “Mixed” model (2PL+GRM) had lower information criterion measures, suggested it has a **better model-data fit** compared to the “Rasch” model (Rasch+PCM).

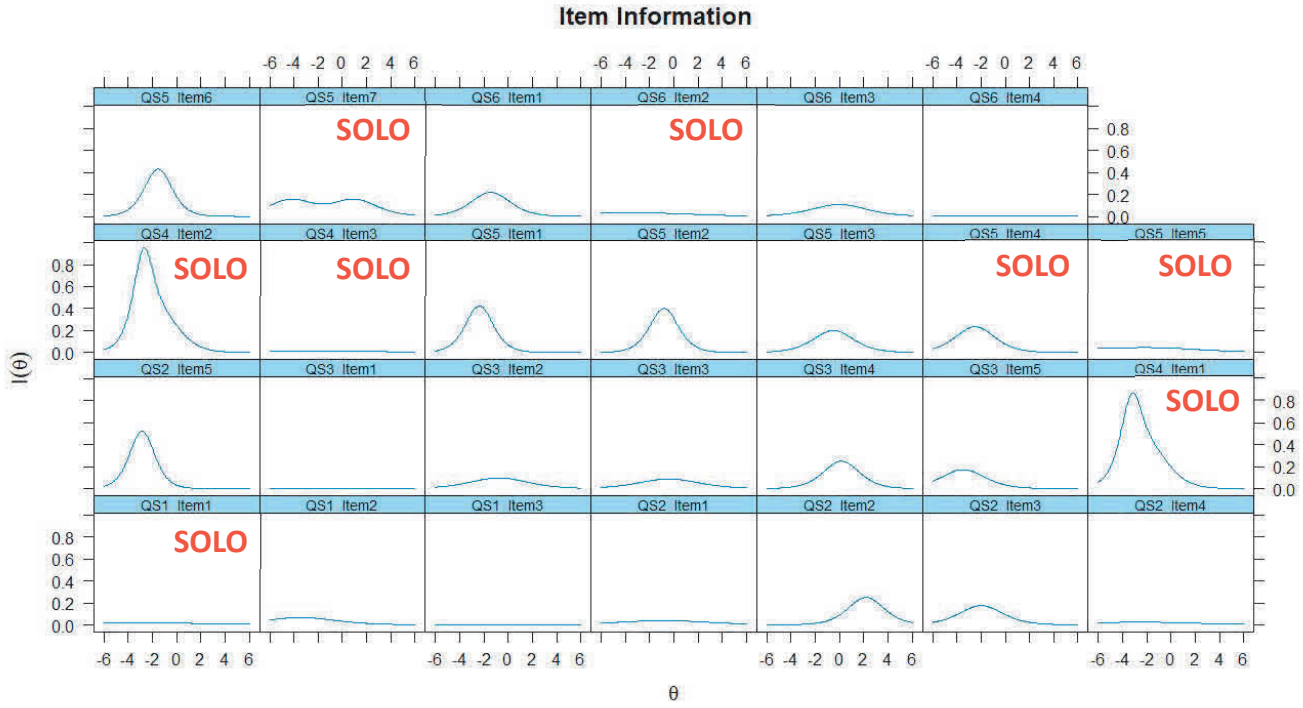
Model Fit (Item Level)

Item	Item Type	Mixed Model				Rasch				Evaluation
		S_X2	df.S_X2	RMSEA.S_X2	p.S_X2	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2	
QS1_Item1	SOLO	55.13506	42	0.01410026	0.084209	94.69308	37	0.03148449	5.92E-07	Bad fit
QS1_Item2	SOLO	31.77808	19	0.0206772	0.0331	31.40175	19	0.02037044	0.036456	Bad fit
QS1_Item3	SOLO	14.92714	20	0	0.780561	28.68487	19	0.01800138	0.071106	M > R
QS2_Item1	MC	10.8239	19	0	0.92962	10.9812	19	0	0.924469	M > R
QS2_Item2	MC	19.44902	18	0.00715381	0.364677	122.1998	15	0.06740422	7.07E-19	M < R
QS2_Item3	MC	12.93205	18	0	0.795594	26.4608	19	0.0157998	0.117853	M > R
QS2_Item4	MC	13.4733	19	0	0.813485	14.58981	19	0	0.748287	M > R
QS2_Item5	MC	16.34973	16	0.00372773	0.42883	33.22923	14	0.02954967	0.00267	M > R
QS3_Item1	MC	26.50447	20	0.01437893	0.14979	52.75184	18	0.03503389	2.87E-05	M > R
QS3_Item2	MC	12.89189	18	0	0.797955	17.50269	20	0	0.620132	M > R
QS3_Item3	MC	15.41399	19	0	0.695947	20.59023	19	0.00729438	0.359878	M > R
QS3_Item4	MC	11.94166	19	0	0.888104	11.41323	19	0	0.909154	M < R
QS3_Item5	MC	9.604437	17	0	0.919374	14.42825	17	0	0.63659	M > R
QS4_Item1	SOLO	35.33388	29	0.01178342	0.193785	33.33589	30	0.00840777	0.308152	M < R
QS4_Item2	SOLO	34.71702	39	0	0.665547	28.85953	38	0	0.857364	M < R
QS4_Item3	SOLO	104.7879	48	0.02742478	4.14E-06	105.6532	46	0.02871268	1.36E-06	Bad fit
QS5_Item1	MC	20.26914	17	0.01105676	0.260695	36.6841	17	0.0271312	0.003713	M > R
QS5_Item2	MC	7.780672	16	0	0.955129	51.28365	19	0.03286625	8.46E-05	M > R
QS5_Item3	MC	15.30866	17	0	0.57326	36.57825	20	0.02295567	0.013141	M > R
QS5_Item4	SOLO	21.12506	26	0	0.73548	25.29667	27	0	0.55784	M > R
QS5_Item5	SOLO	55.37055	41	0.01492727	0.066302	83.31112	40	0.02623647	7.06E-05	M > R
QS5_Item6	MC	13.68145	17	0	0.689509	57.81562	19	0.03603812	8.53E-06	M > R
QS5_Item7	SOLO	37.09606	32	0.01006185	0.245591	58.42816	31	0.02371664	0.002058	M > R
QS6_Item1	MC	30.986	18	0.02141594	0.028895	49.17352	19	0.03177399	0.000173	Bad fit
QS6_Item2	SOLO	49.34036	36	0.01534858	0.068377	78.70634	35	0.02817564	3.29E-05	M > R
QS6_Item3	MC	18.70031	18	0.0049733	0.410492	26.9658	19	0.01632576	0.105452	M > R
QS6_Item4	MC	20.30359	18	0.00901991	0.31599	34.41642	17	0.02552058	0.007415	M > R

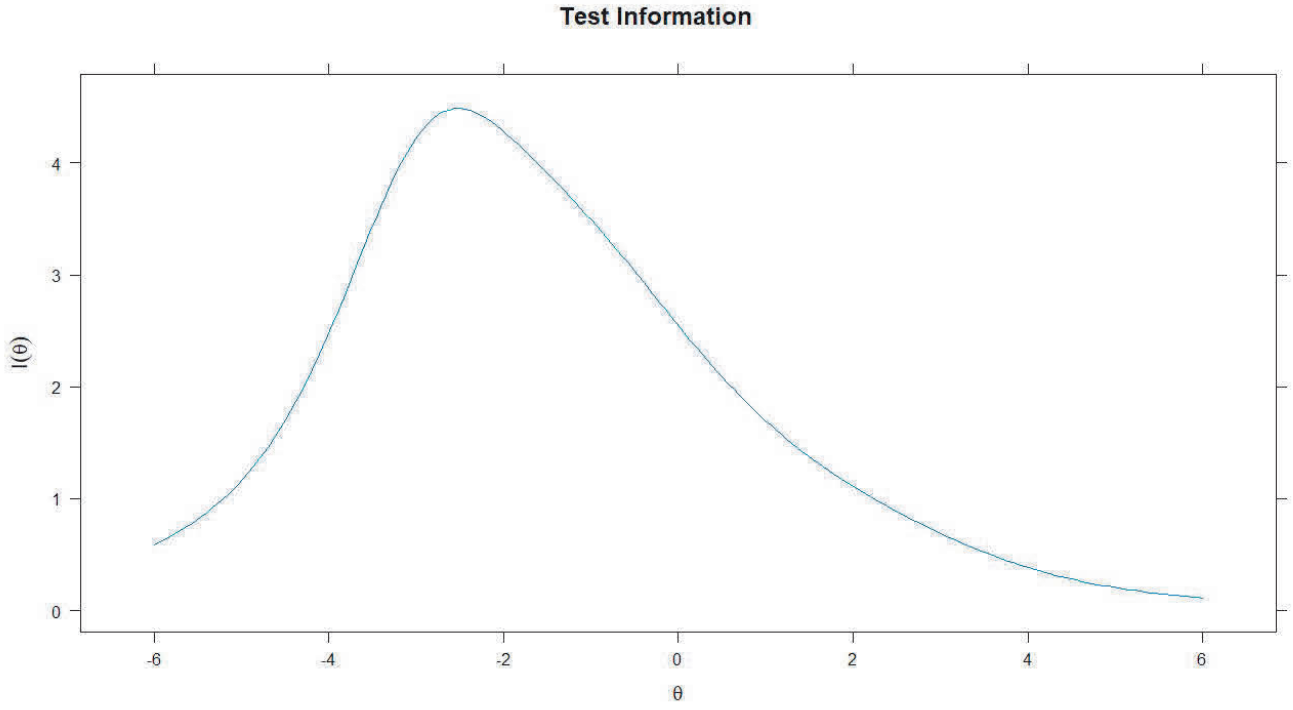
IRT Results



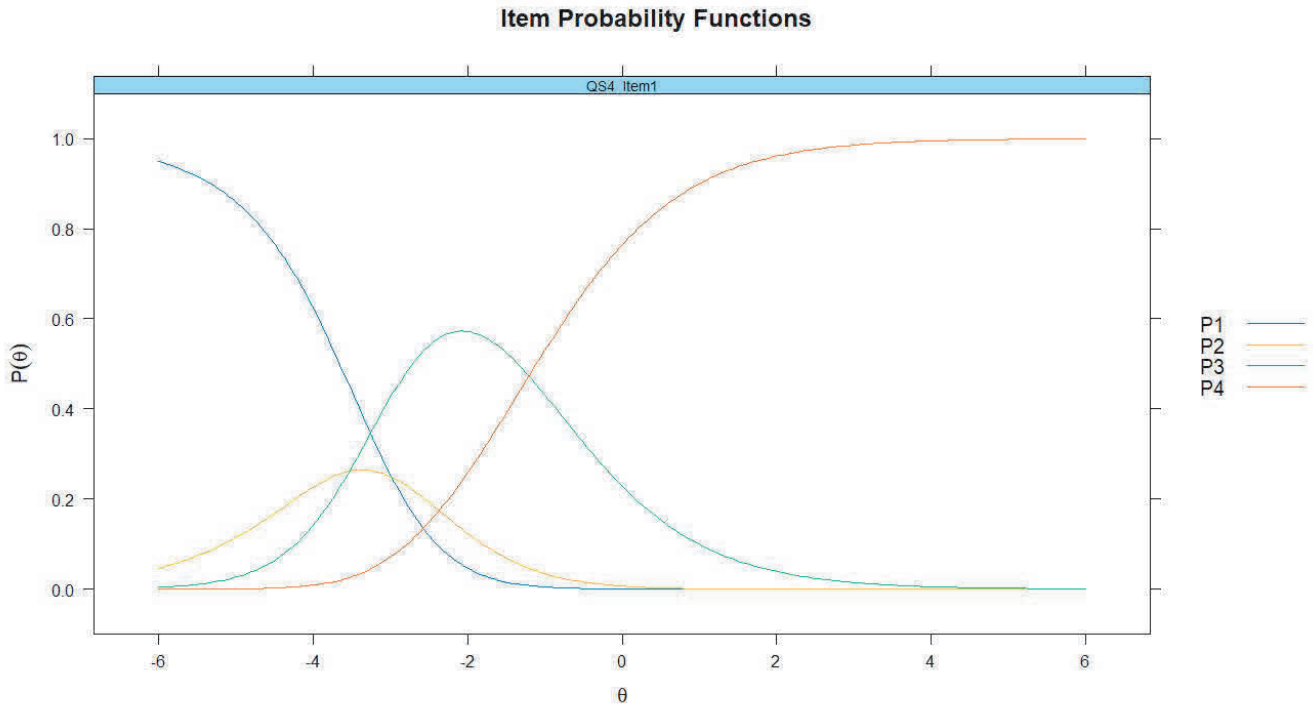
IRT Results



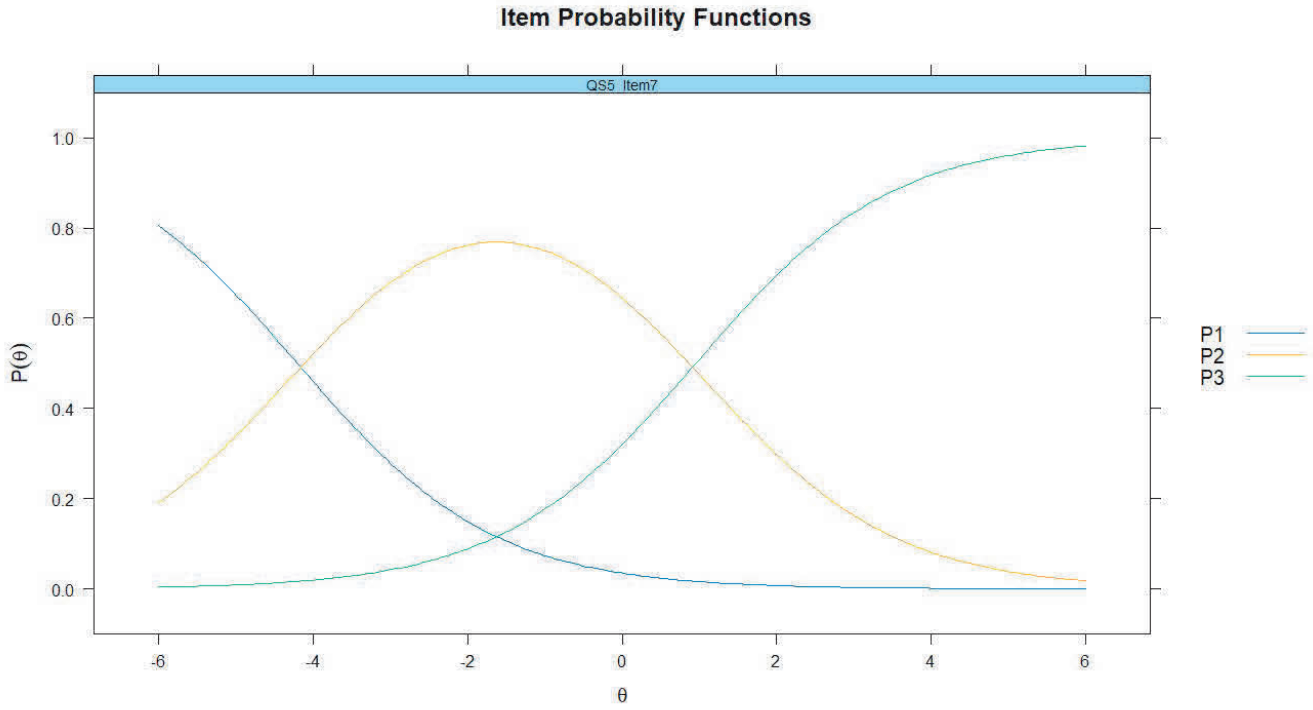
IRT Results



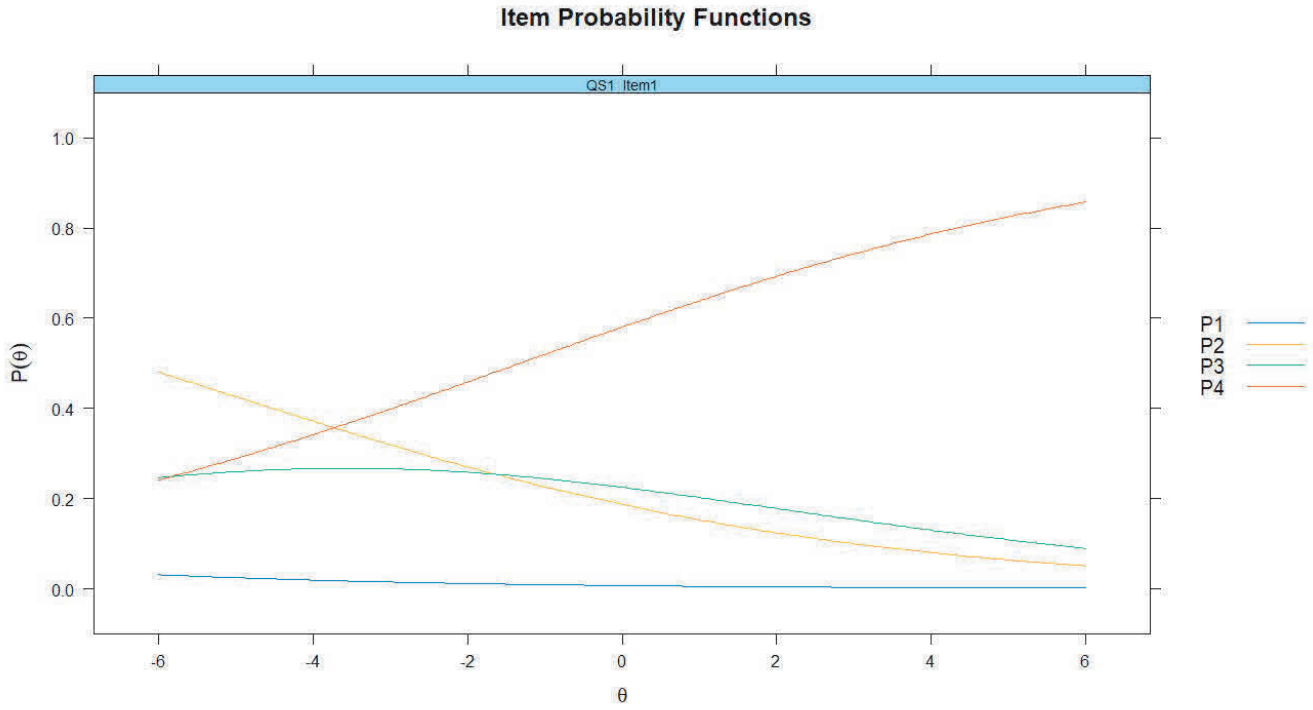
Category Response Curve QS4_Item1



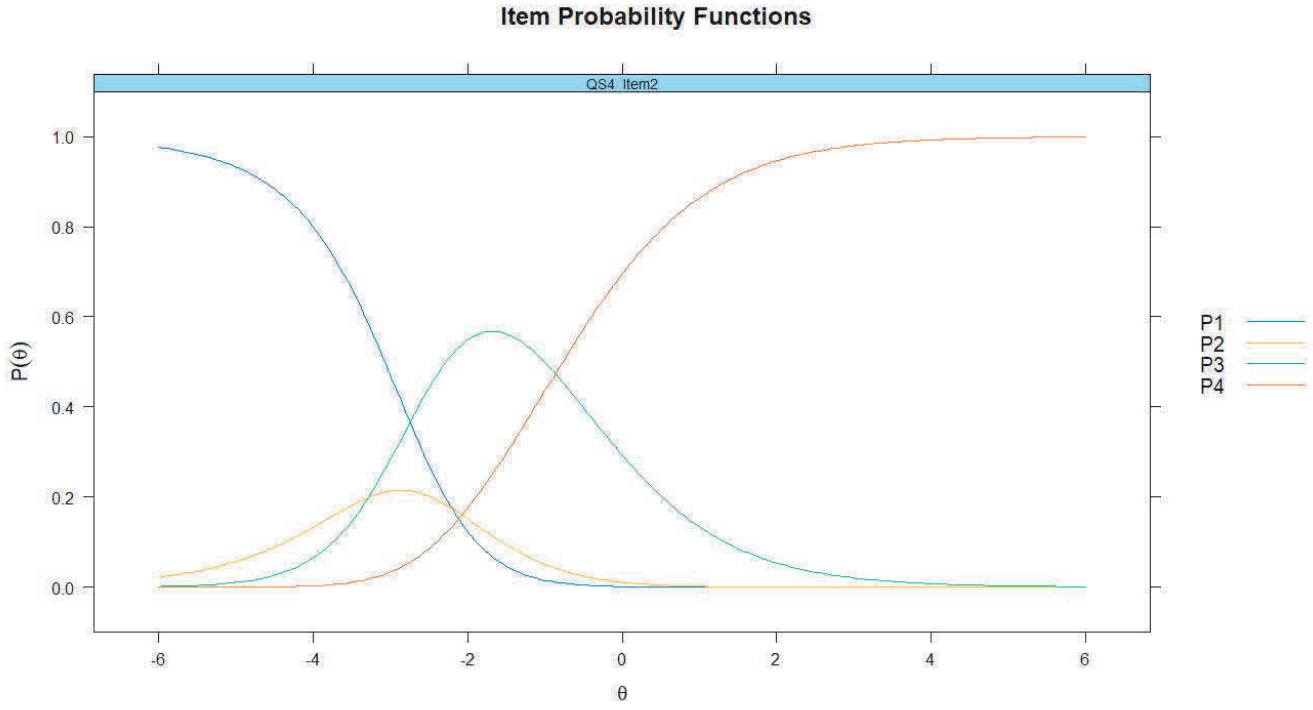
Category Response Curve QS5_Item7



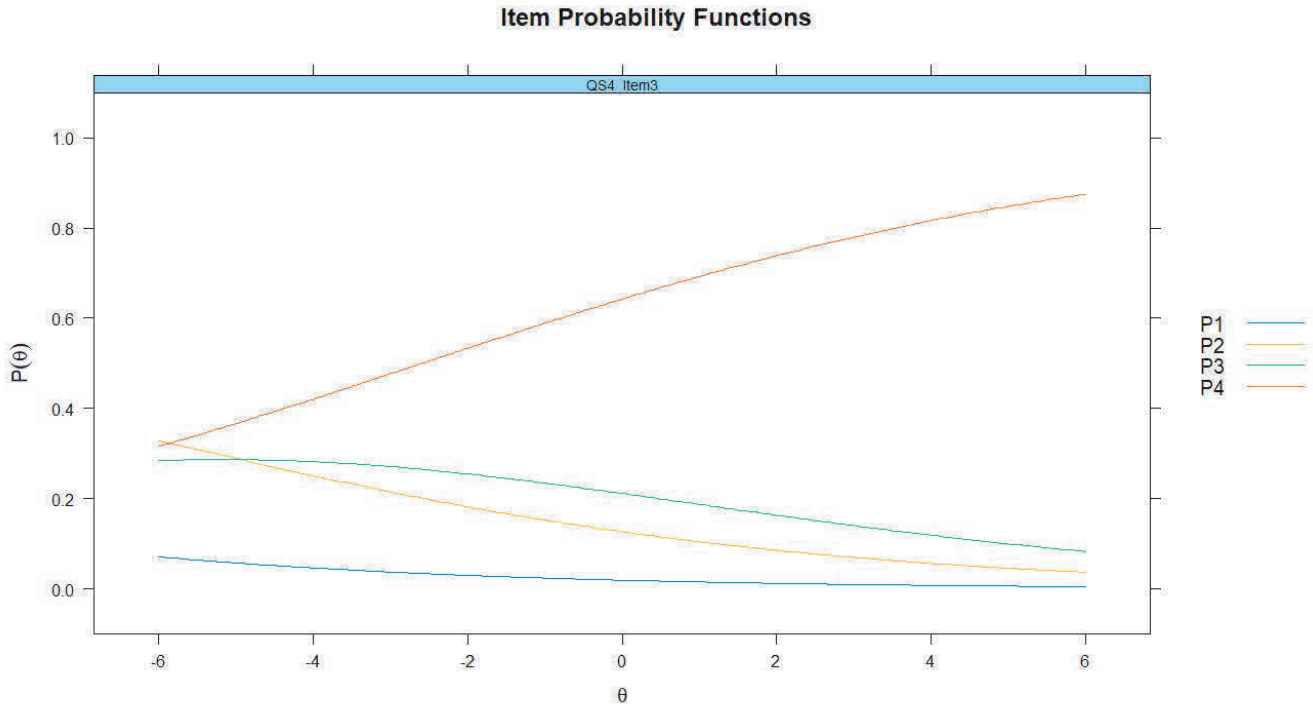
Category Response Curve QS1_Item1



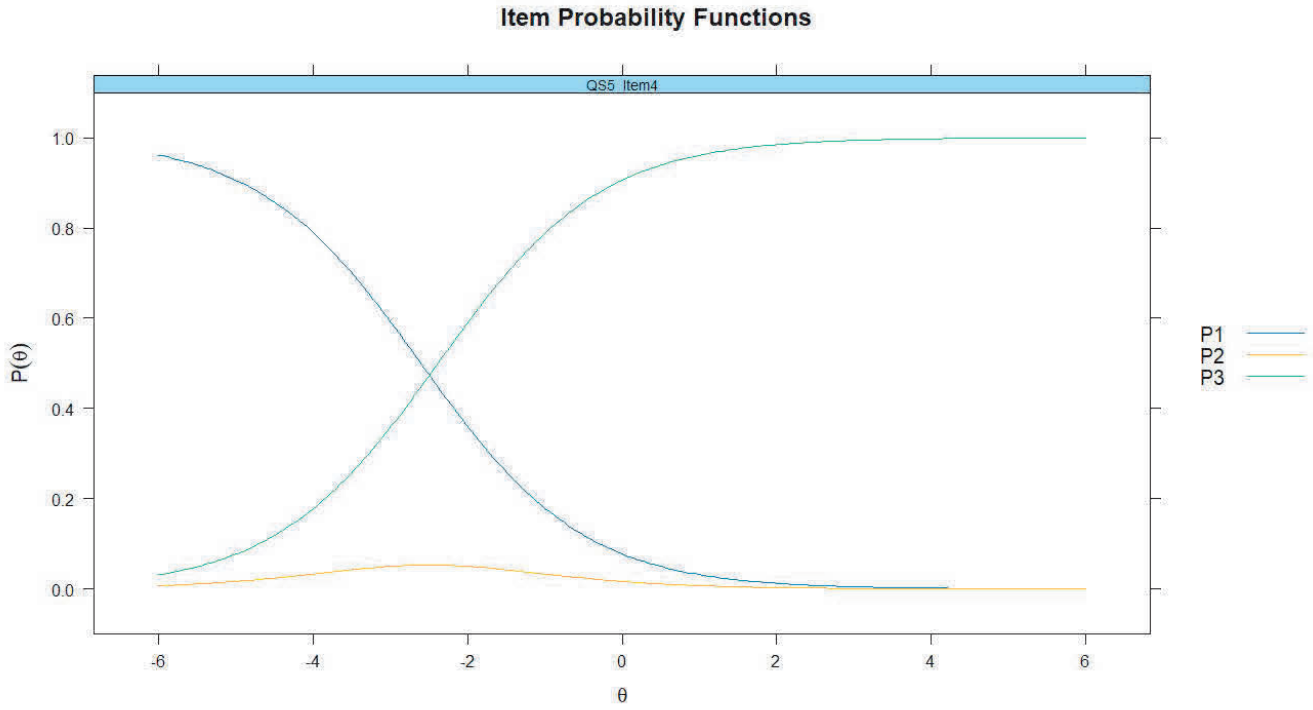
Category Response Curve QS4_Item2



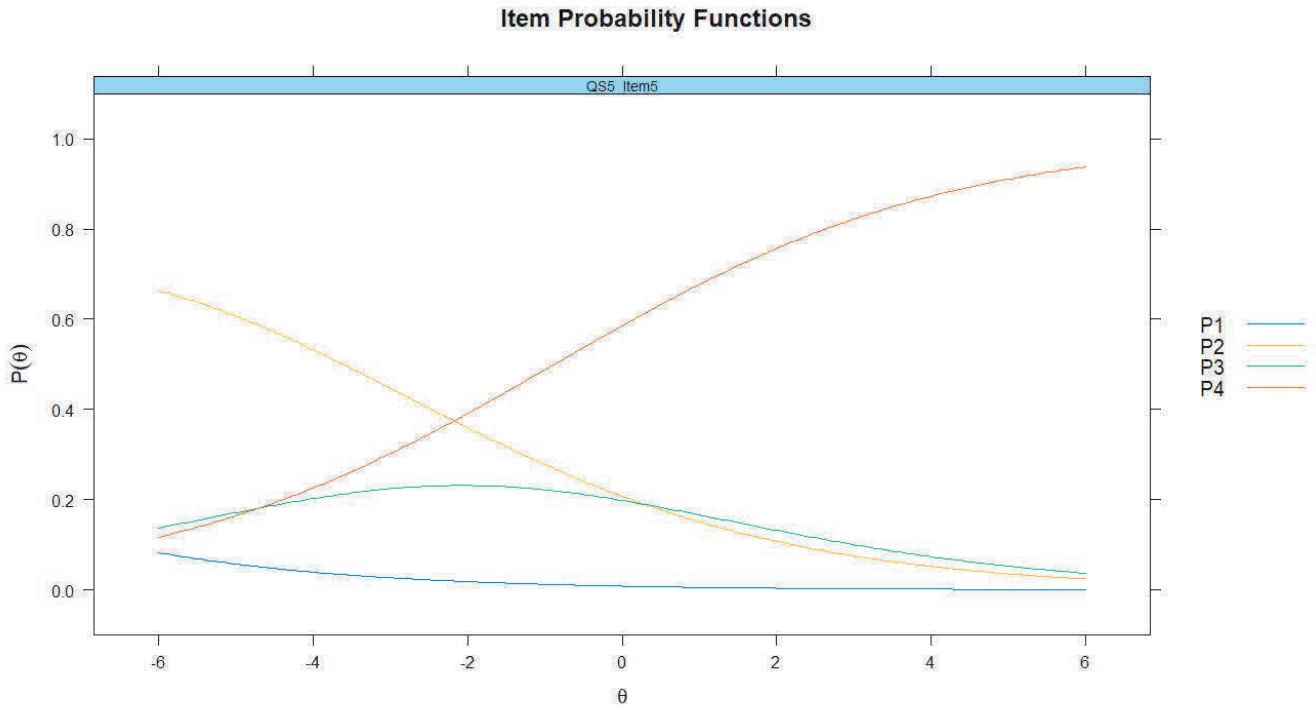
Category Response Curve QS4_Item3



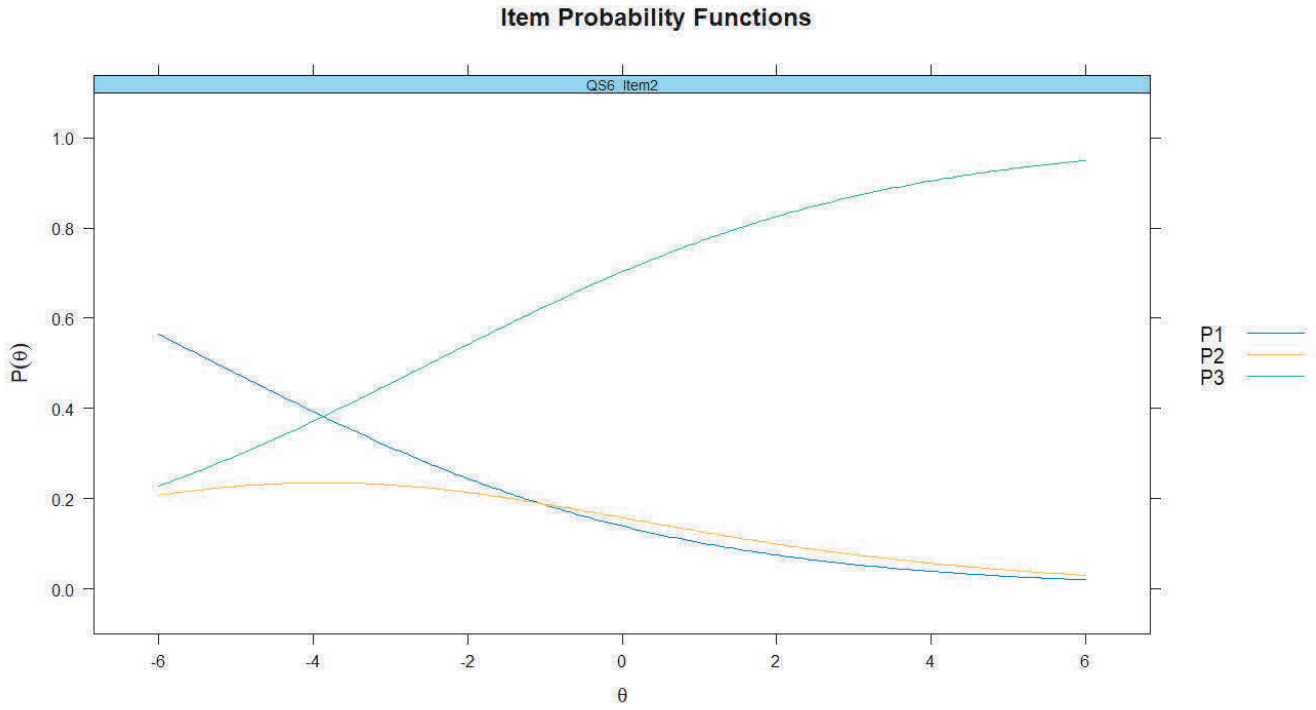
Category Response Curve QS5_Item14



Category Response Curve QS5_Item5

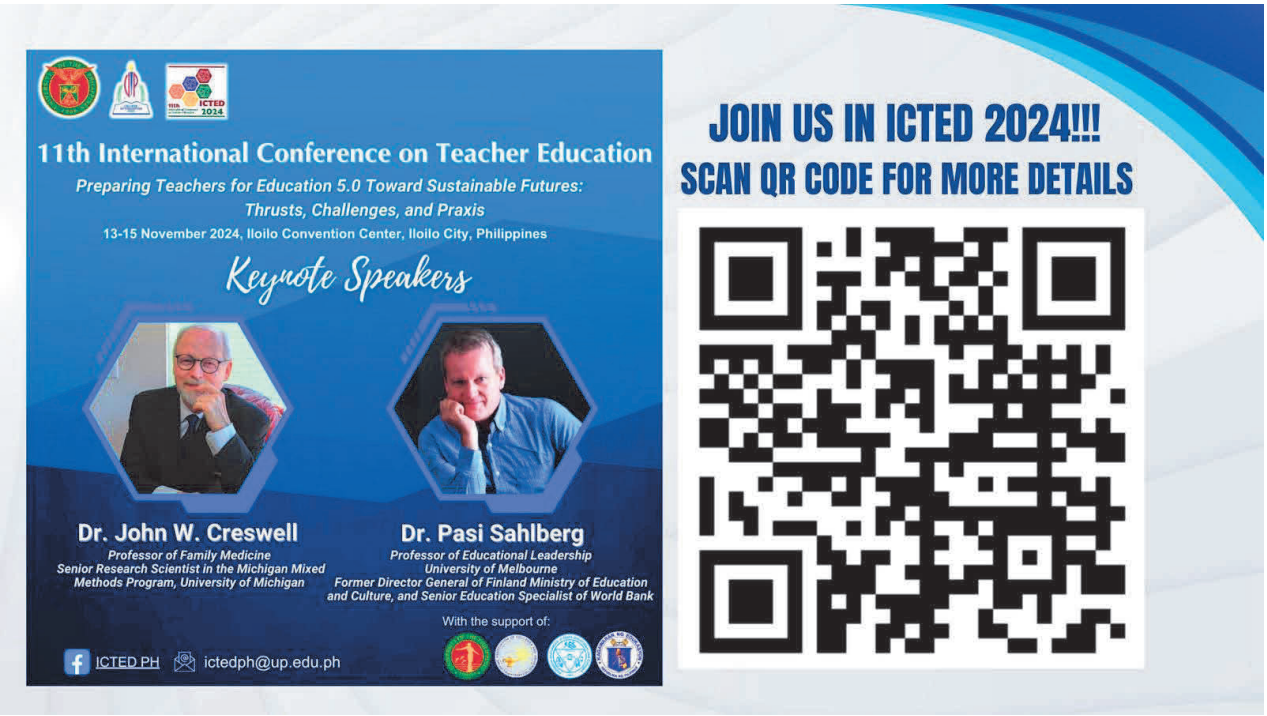


Category Response Curve QS6_Item2



Discussion

- The two-parameter logistic (2PL) model and the Graded response model (GRM) fit better with the given data than the Rasch and Partial Credit model.
- The hypothesized categories in SOLO items **need to be empirically validated** to see if these are functioning.
- Caution is recommended in using it for grading purposes especially if the SOLO items are not empirically validated.
- The type of item used has implications in computing the total score. We recommend that **IRT should be used in generating test scores** for mixed item format tests.



ACTRC

lpcagasan@up.edu.ph
l.cagasan@actrc.org

www.actrc.org
www.facebook.com/ACTRC.org