ACTRC

On the Usefulness of Testlet Item Response Theory Models: An Illustration Using a Reading Assessment

Karizza Bianca Loberiza
Kevin Carl P. Santos, PhD

National Conference on Educational Measurement and Evaluation 2024
August 30, 2024

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

Australian Aid

Suggested Citation

Loberiza, K., & Santos, K. (2024, August 29-31). On the Usefulness of Testlet Item Response Theory Models: An Illustration Using a Reading Assessment [Conference panel presentation]. National Conference on Educational Measurement and Evaluation 2024, Manila, Philippines.

Aim of the Presentation

- To compare the application of a traditional item response theory (IRT) model with an IRT testlet model to the same reading assessment
- To examine the impact of ignoring the “testlet effect” in educational assessments

What are testlets?

- Testlets (also called *test bundle*) are set of items that share a common stimulus such as graph, table, diagram, map, item stems, and scenario (Wang et al., 2005).
- Testlets have been used in education and psychological test.
- Its benefits include reducing testing time and cost, and minimizing fatigue among test takers.

Local Independence

- An essential assumption of traditional IRT models is **local independence.**
- Items are assumed to be independent when the ability level has been partialed out.
- When there are sets of items that share the same stimulus, this assumption is **not** satisfied.
- Standard IRT models may not work properly with testlets, resulting in distortion of parameter and ability estimates.

Fitting a Standard IRT Model to Testlets

About the Test	
Area	Reading Comprehension
Grade Level	Grade 5
Number of Items	49
Item Type	Multiple-Choice
Number of testlets	9
Number of test-takers	1570

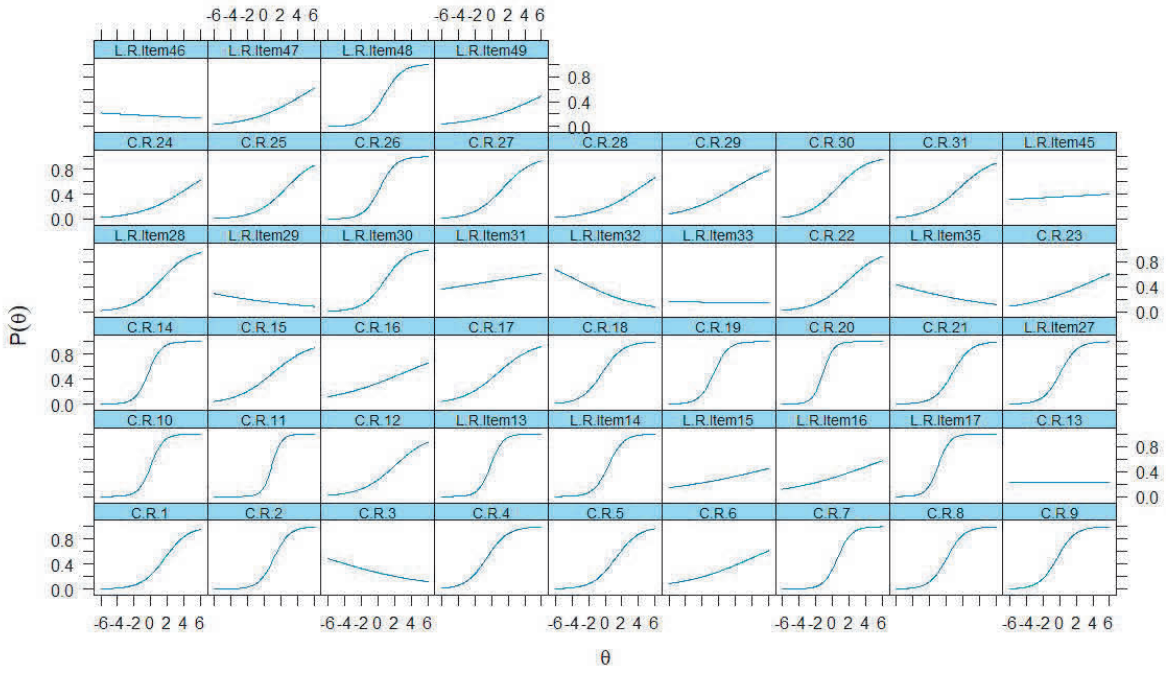
Fitting the 2PL IRT Model

- The 2-parameter logistic model was fitted to the data

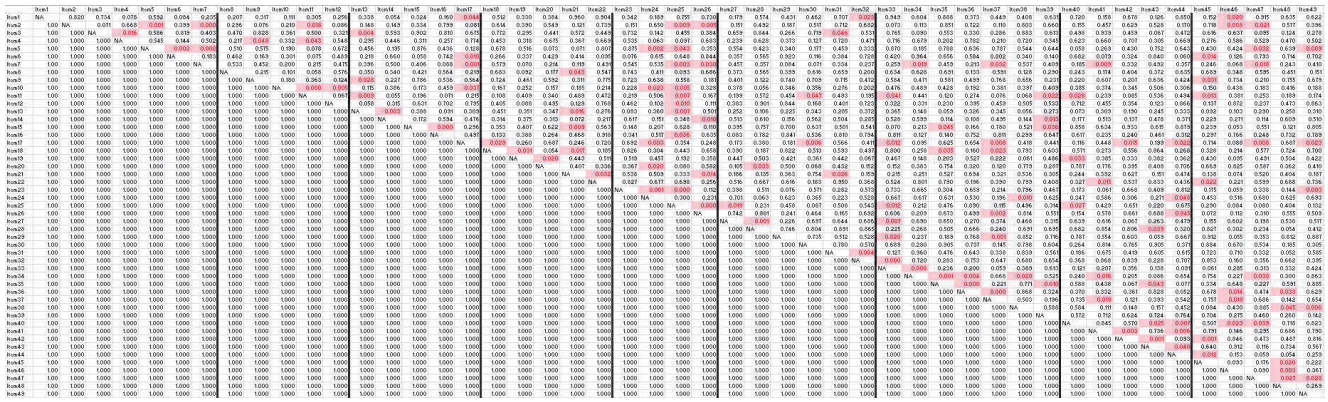
$$P\left(X_{ij}=1|\theta_j\right)=\frac{1}{1+e^{-a_i\left(\theta_j-b_i\right)}}$$

where θ_j is the ability of examinee j; and a_i, b_i are, respectively, the discrimination and difficulty parameters to item i.

Fitting the 2PL IRT Model



Test for Local Independence (Chen and Thissen, 1997)



Test for Local Independence (Chen and Thissen, 1997)

Testlet 1

	Item1	Item2	Item3	Item4	Item5	Item6	Item7
Item1	NA	0.820	0.734	0.078	0.592	0.084	0.235
Item2		NA	0.071	0.668	0.001	0.399	0.000
Item3			NA	0.016	0.586	0.819	0.403
Item4				NA	0.545	0.144	0.502
Item5					NA	0.002	0.002
Item6						NA	0.183
Item7							NA

Fitting the 2PNO-testlet model

- The testlet response model for person *p* at item *i* is defined as

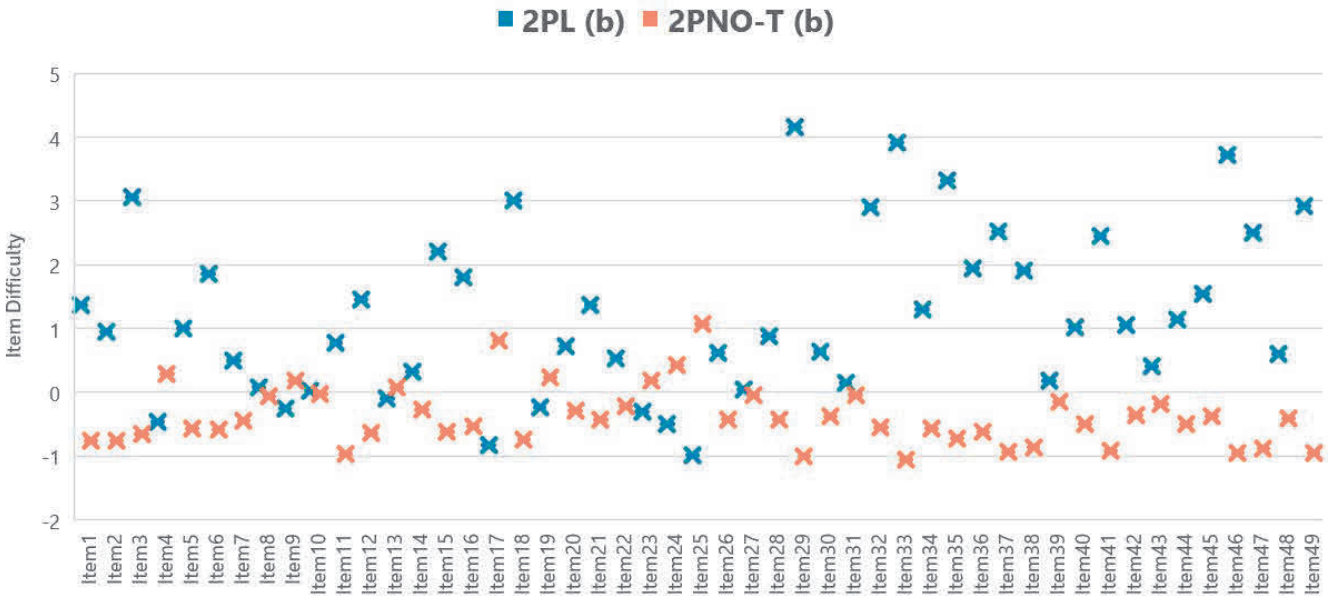
$$P(X_{pi}=1)=c_i+(1-c_i)\Phi(a_i\theta_p+\gamma_{p,t(i)}+b_i) \quad , \quad \theta_p\sim N(0,1), \gamma_{p,t(i)}\sim N(0,\sigma_t^2)$$

For two-parameter normal ogive (2PNO)-testlet model, guessing is set to zero

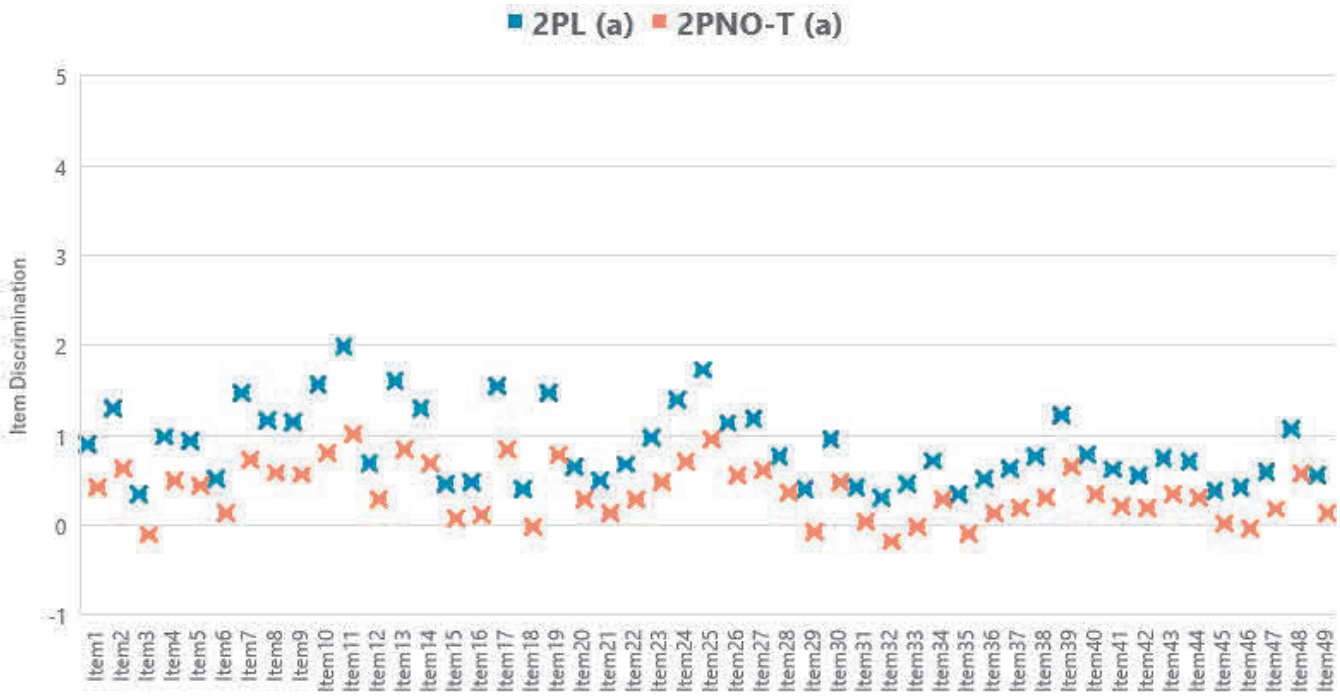
After fitting the testlet model, marginal item parameters are calculated (integrating out testlet effects) according the defining response equation

$$P(X_{pi}=1)=c_i+(1-c_i)\Phi(a_i^*\theta_p+b_i^*)$$

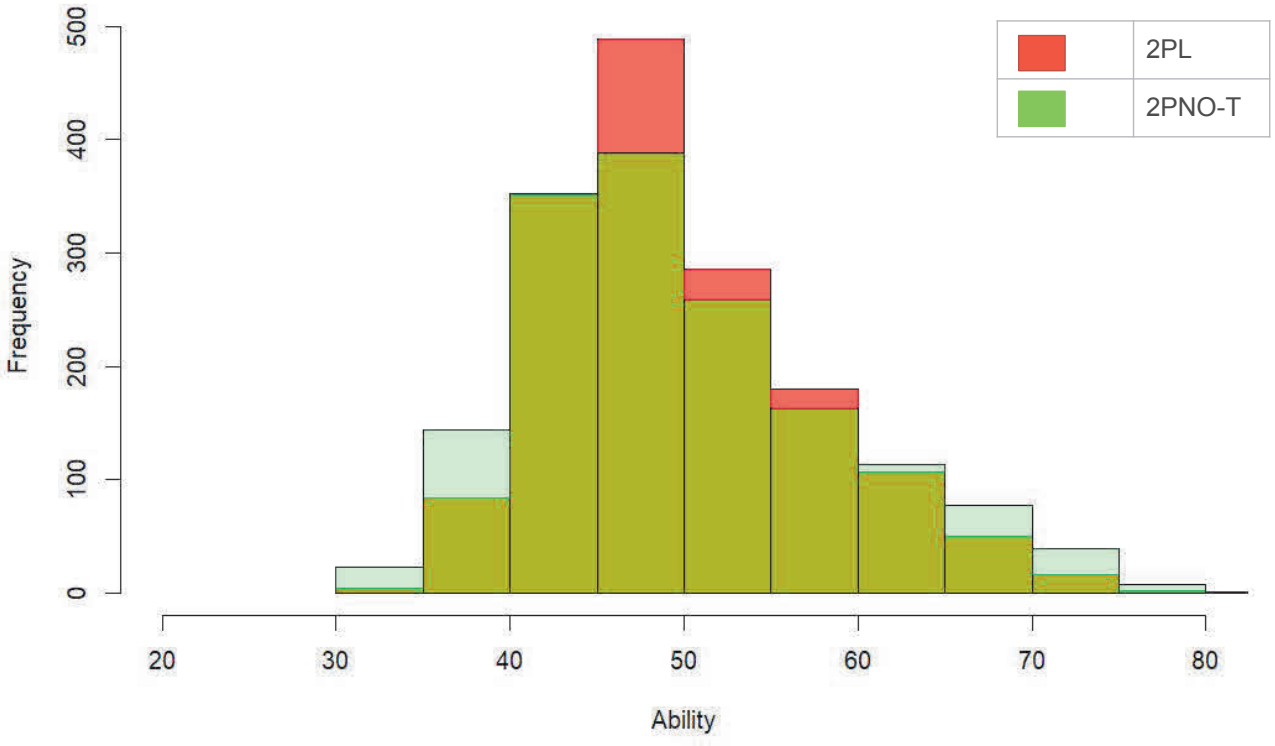
Comparison of Item Difficulty Using 2PL and 2PNO-Testlet Models



Comparison of Item Discrimination Using 2PL and 2PNO-Testlet Models



Comparison of Ability Estimates



Implications of fitting traditional IRT model to a testlet response

- Despite having several procedures for estimating testlet response, the dependency is often ignored in practice, and standard IRT models are used instead (Paap & Veldkamp, 2012).

- Ignoring the possible dependence between the items within a testlet leads to an **overestimation of item parameters and ability estimates.**

- These biased estimations lead to inaccurate inferences about the parameters.

- Hence, the possibility of local dependence should be considered when calibrating items from a testlet response.

ACTRC

k.loberiza@actrc.org

www.actrc.org

www.facebook.com/ACTRC.org

@ACTRC_edu