

Psychometric Calibration of an Eighth-Grade Reading Comprehension Test

Weon Kim and Amy Elleman
Literacy Studies Ph. D. Program

Abstract

The purpose of this study is to conduct item and test analyses for a large data set for a reading comprehension test. Both classical test theory (CTT) and item response theory (IRT) were used to evaluate psychometric properties of a 31-item eighth-grade reading comprehension test.

The results showed a good test reliability index in CTT and a strong model-data fit for the 3-parameter logistic model (3-plm) in IRT.

For model comparisons, likelihood ratio tests, Bayesian information criterion (BIC), and Akaike information criterion (AIC) were applied. Person's ability (θ -parameter), item discrimination index (a -parameter), item difficulty index (b -parameter) and pseudo-chance parameter (c -parameter) were estimated through the expected-a-posteriori (EAP) estimation method. Item information functions, test information function (TIF), and item characteristic curves (ICCs) were also reported.

Introduction

Reading comprehension is an extraordinarily complex task. The complexity of reading comprehension assessment mirrors the complexity of reading comprehension.

This complicated area of reading comprehension assessment is directly related to the area of psychometrics concerning mental measurement. In psychometrics, there are two major theories—CTT and IRT. CTT has been the most popular test theory until the turn of the 21st century.

IRT is a new alternative way for mental measurement. IRT has many theoretical advantages over CTT. First, it has falsifiable models while CTT is a tautology. Second, IRT has invariant person and item indices whereas CTT has variant indices.

Third, IRT gives more accurate parameter estimations than CTT. Fourth, IRT is an individualized test (e.g., Graduate Record Examination). Fifth, fewer items are needed to estimate examinees' ability.

It is clear that IRT has many theoretical and practical advantages over CTT to evaluate examinees' mental process and behaviors such as reading comprehension.

The purpose of this study is to conduct item and test analyses and to evaluate psychometric properties using a large data set with 31 items of an eighth-grade reading comprehension test based on both CTT and IRT.

Method

The data set was obtained from a testing company. The test was administered in May, 2012/2013, in 14 states in the United States.

The test was administered online. The test followed the Common Core State Standards: literature, information, language, and writing

- Participants: 9,763 eighth-grade students
- Proportion of gender: similar for both female and male.
- Majority population: Caucasian students
- Minority populations: African American (17%), Asian (3.5%), and American Indians (2.3%)
- English as a Second Language (ESL) students: approximately 7.8% of the population

The XCalibre software from the Assessment Systems Corporation was utilized to analyze the test and items with the given data based on both CTT and IRT.

Figure 1. Examples of poor-fit of the ICC to the data (item 10 and item 11).

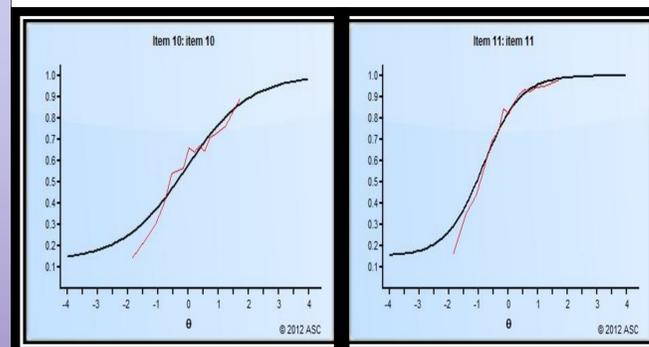
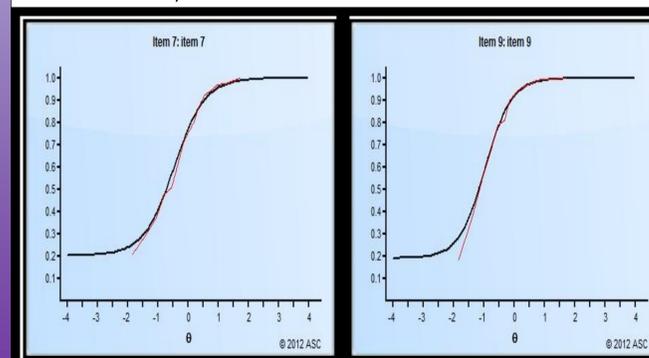


Figure 2. Examples of good-fit of the ICC to the data (item 7 and item 9).



Results

CTT analyses

- Cronbach's alpha = .89, Mean = 19.36 (total 31 items), SD = 7.01

IRT analyses

- A well-fitting 3-plm with stable parameter estimates based on likelihood ratio test, BIC, and AIC
- Means of θ -parameter, a -parameter, b -parameter, and c -parameter: appropriate ranges for the test
- Items 10 and 11: poor-fits to the given model based on the item statistics (Figure 1)
- Item 7 and item 9: examples of good-fit of the ICC to the data (Figure 2)
- Maximum item information value of item 30: the highest among the other items' maximum information values, Maximum item information value of item 22: the least one among 29 items (Figure 3)
- Test information function; when theta was -0.25, the maximum value of TIF was 12.2568; this test would be more useful for assessing abilities of examinees whose theta value is -0.25 (Figure 4)

Figure 3. Examples of good items with the most values (items 18 and 30) and poor items with the least values of item information function (items 3 and 22).

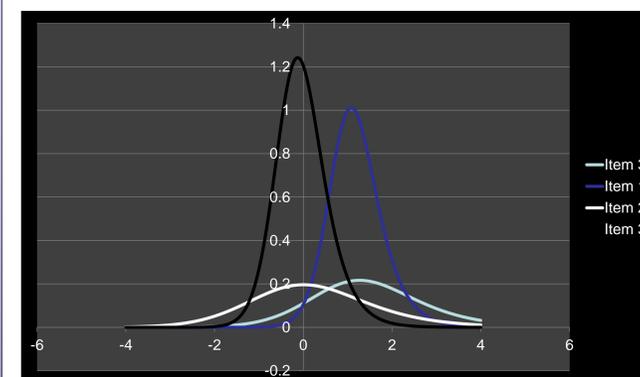
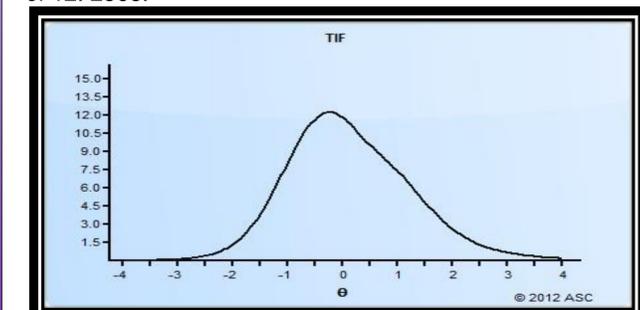


Figure 4. Test information function with the maximum value of 12.2568.



Discussion

This study applied a new psychometric theory, IRT, along with the traditional CTT. The results showed a good test reliability index of .89 from Cronbach's alpha and a strong model-data fit for the 3-plm in IRT with reasonable item parameters. This reading comprehension test contained solid psychometric characteristics in both CTT and IRT.

As Keenan et al. (2008) address, examinees' ability for reading comprehension changes from test to test. Item indices change from sample to sample. Reading comprehension results from different tests are neither comparable nor interchangeable.

IRT may be a partial solution for these problems because IRT provides invariant person and item indices provided that the model fits the data. Power for item selection is also a very important advantage of IRT. Items can be evaluated and selected based on the item information functions.

The structure of reading comprehension tests is unique; items are correlated with each other because the cluster of items comprises questions related to the same passage. Thus, the conditional independence assumption of IRT may be violated. The testlet response theory (TRT) model is needed when the conditional independence assumption of IRT is violated.

For future study, TRT models will be applied to the data and characteristics of each TRT model will be evaluated.

References

- Betjemann, R. S., Keenan, J. M., Olson, R. K., & DeFries, J. C. (2011). Choice of reading comprehension test influences the outcomes of genetic analyses. *Scientific Studies of Reading, 15* (4), 363-382.
- Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U.A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 102* (3), 687-700.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*, 2. Newbury Park, CA: Sage.
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. New York, NY: Cambridge University Press.