

INTRODUCTION

Research Question:

- What is the strength and weakness of each item in CERT?

Background Information:

- No Child Left Behind (NCLB, 2001)*: Need for equitable testing of all students
- Educators and testing companies are charged with the task of creating tests that will measure academic growth and assess student readiness for these high stake tests at all levels
- Classical test theory (CTT) and item response theory (IRT) have both been applied to psychometrically validate education tests
- The benefits of IRT is that it can separate out different characteristics between examinees and tests (Thomas, 2011)
- IRT looks at all components of the system including the underlying variables (latent) that may contribute to the results but are not easily measured (Thomas, 2011)
- Kim and Nicewander (1993) point out that ability scores do not change with respect to the difficulty of the test whereas scores on a conventional test would vary with the difficulty of the test items
- The most important difference between CTT and IRT is that in CTT uses a common estimate that is assumed to be equal for all individuals irrespective of their attribute levels. (Jabrayilov, Emons, & Sijtsma, 2016)
- In IRT the measurement precision depends on latent-attribute value. (Jabrayilov, Emons, & Sijtsma, 2016)
- Researchers point out that the reliability of the ability estimates were essentially better than the number-right score, even though they were biased in both extremes of the distribution on the moderate test and “opposite extremes for Difficult and Easy tests”. (Kim & Nicewander, 1993)

Purpose:

- Classical test theory (CTT) and item response theory (IRT) have both been applied to psychometrically validate the College Equipped Readiness Test (CERT)
- The implications of the study were to decompose the test and utilize the IRT analysis to select the best test items that would predict academic performance in each domain and that would ultimately guide teacher instruction.

METHODS

Participants:

- 123 magnet high school students in the Middle Tennessee area

Measures:

- CERT, like ACT tests, contain four multiple-choice subject tests: English, Mathematics, Reading, and Science
- The English test contained 75 total items, Math 60-items, Reading 40-items and Science contained 40-items

Procedures:

- All items were calibrated using Xcalibre
- Xcalibre uses the expectation-maximization approach to calibrate item parameters
- Both CTT item indices (p -value and item-test correlation) and estimated IRT parameters (a - and b -parameters) were compared for similarities and dissimilarities
- The 1-parameter (1PLM), 2-parameter (2PLM), and 3-parameter (3PLM) logistic models were run and the best-fitting model was selected through the delta chi-square tests
- The 2PLM was chosen as the best model for each subtest.
- The 2PLM predicts the probability of a successful answer using two parameters, difficulty (b) and discrimination (a)
- The steeper the slope, the higher the discrimination of the item and the ability to detect subtle differences in the respondents.

CONCLUSIONS

- With new software such as Xcalibre CTT and IRT analysis that are done by psychometricians in major testing companies can now be performed by educators at the district level
- These analyses could be used to select appropriate testing programs to purchase
- Analysis can guide PLC and Curriculum Lead teams to construct and analyze CFA's

REFERENCES

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (2010). *Fundamentals of item response theory*. Newbury Park: Sage.
- Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Applied Psychological Measurement*, 40(8), 559–572. doi: 10.1177/0146621616666406
- Kim, J. K., & Nicewander, W. A. (1993). Ability estimation for conventional tests. *Psychometrika*, 58(4), 587–599. doi: 10.1007/bf02294829
- Thomas, M. L. (2011). The value of item response theory in clinical assessment: A Review. *Assessment*, 18(3), 291–307. doi: 10.1177/1073191110374797

RESULTS: READING

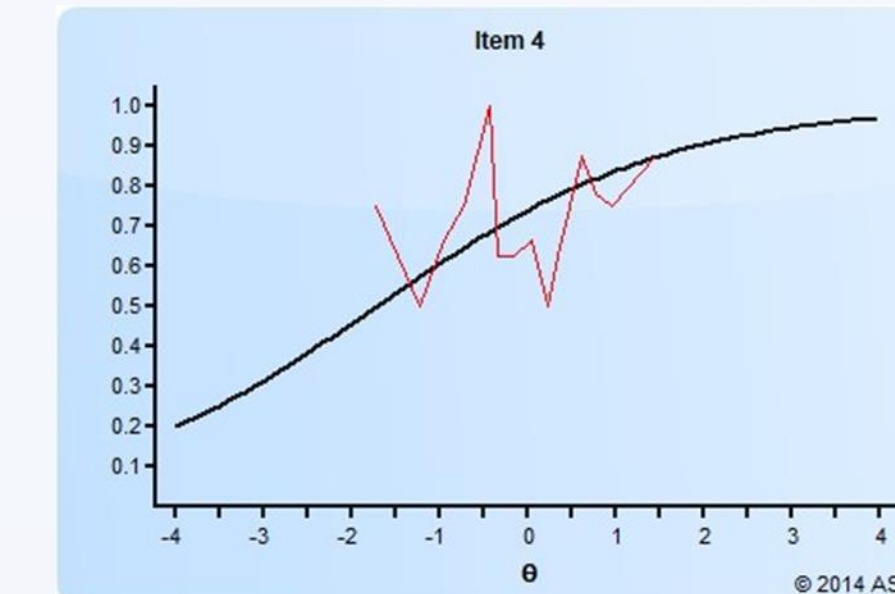
Table 1: Descriptive statistics for CERT Reading test

Item	CTT		IRT 2PLM	
	P	R	a	b
4	0.715	0.039	0.350	-1.672
10	0.577	0.017	0.337	-0.590
26	0.878	0.446	0.751	-2.006
28	0.431	0.478	0.719	0.290

Full Test $M = 28.91$, $SD = 5.10$, Cronbach's $\alpha = .777$

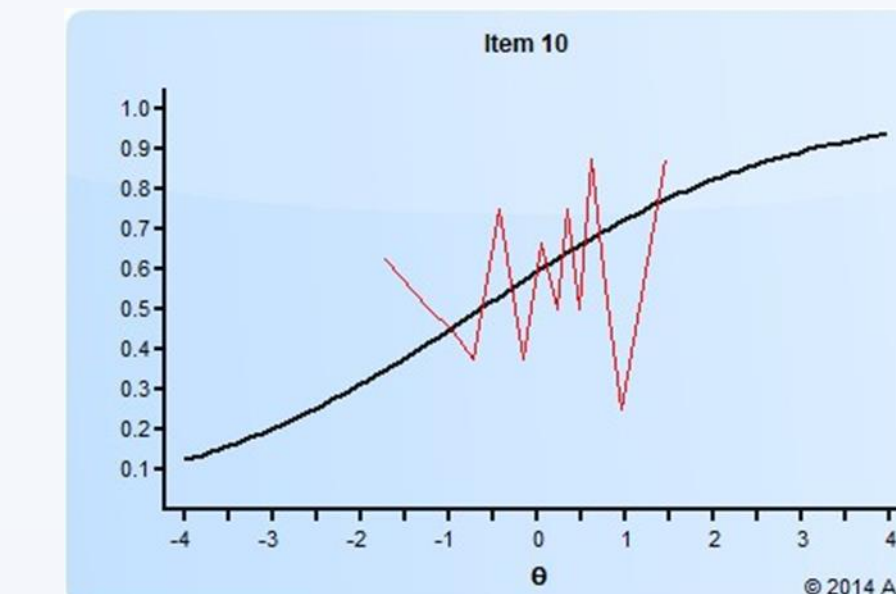
Weak Test Items

Figure 1a: Item Response Function



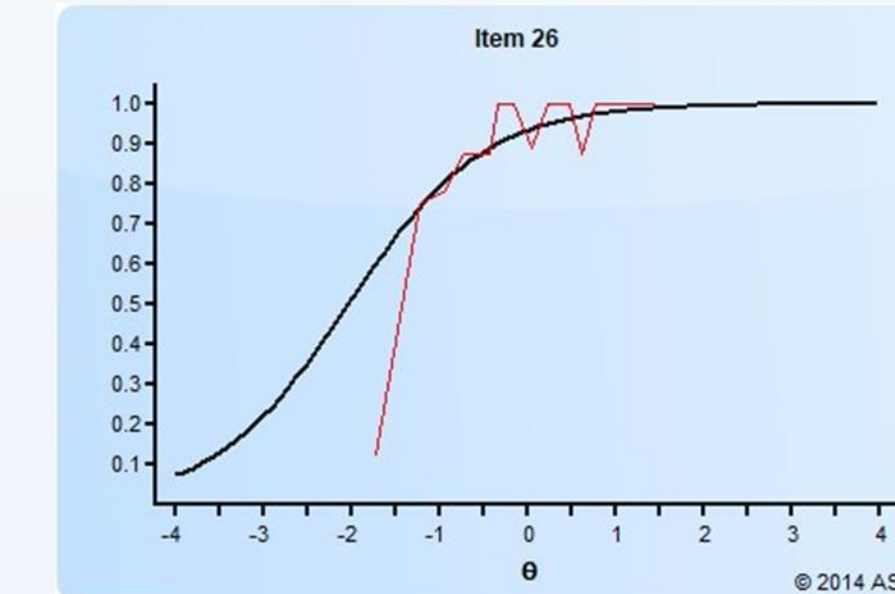
Note: Y-axis = proportion correct units; R Y-axis = number correct units

Figure 1b: Item Response Function



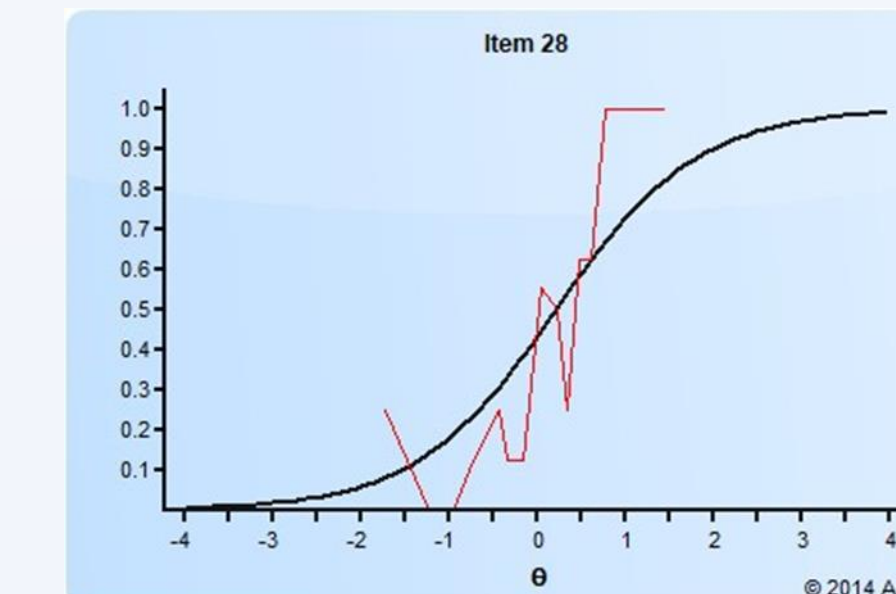
Strong Test Items

Figure 1c: Item Response Function



Note: Y-axis = proportion correct units; R Y-axis = number correct units

Figure 1d: Item Response Function



RESULTS: ENGLISH

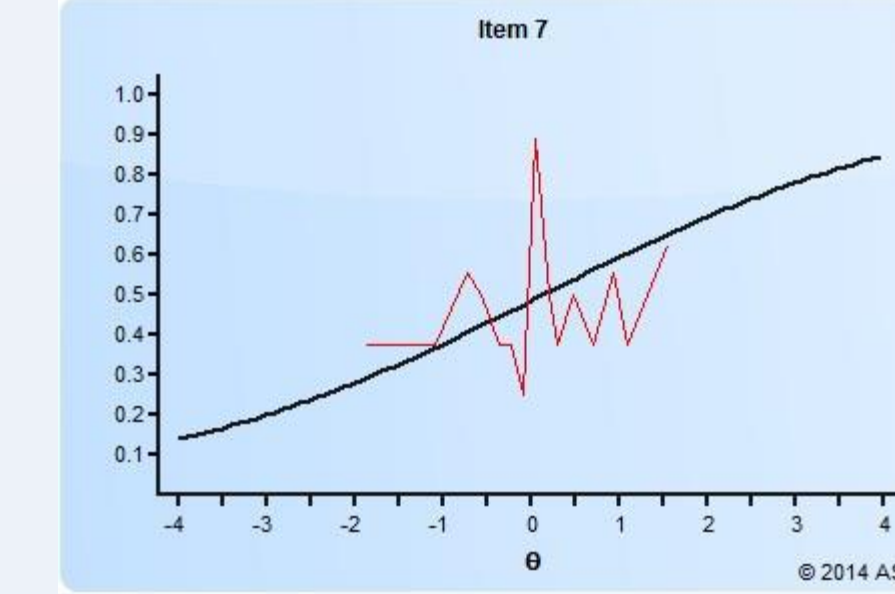
Table 2: Descriptive statistics for CERT English test

Item	CTT		IRT 2PLM	
	P	R	a	b
5	0.854	0.314	0.777	-2.518
7	0.472	-0.021	0.435	0.208
30	0.350	0.257	0.654	0.991
37	0.220	-0.104	0.485	2.500

Full Test $M = 59.75$, $SD = 6.69$, Cronbach's $\alpha = .798$

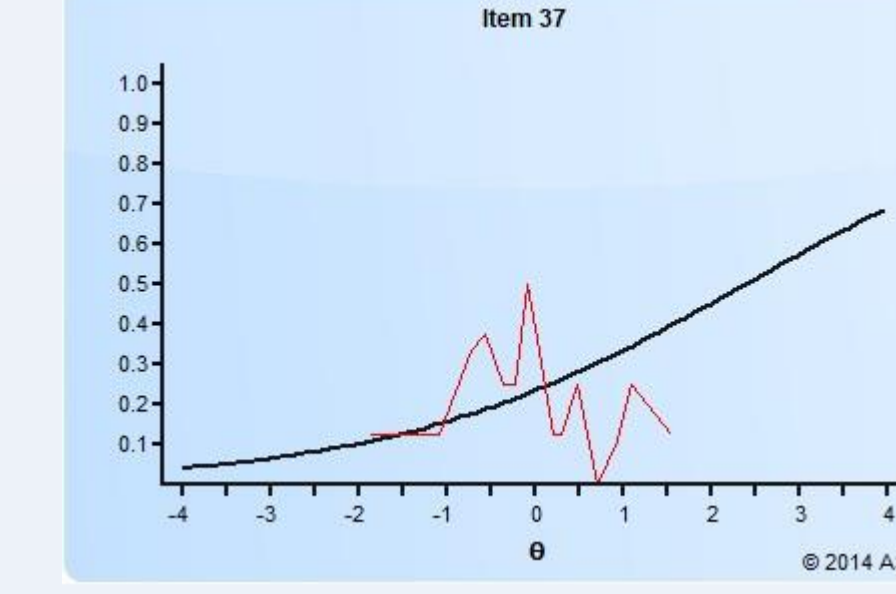
Weak Test Items

Figure 2a: Item Response Function



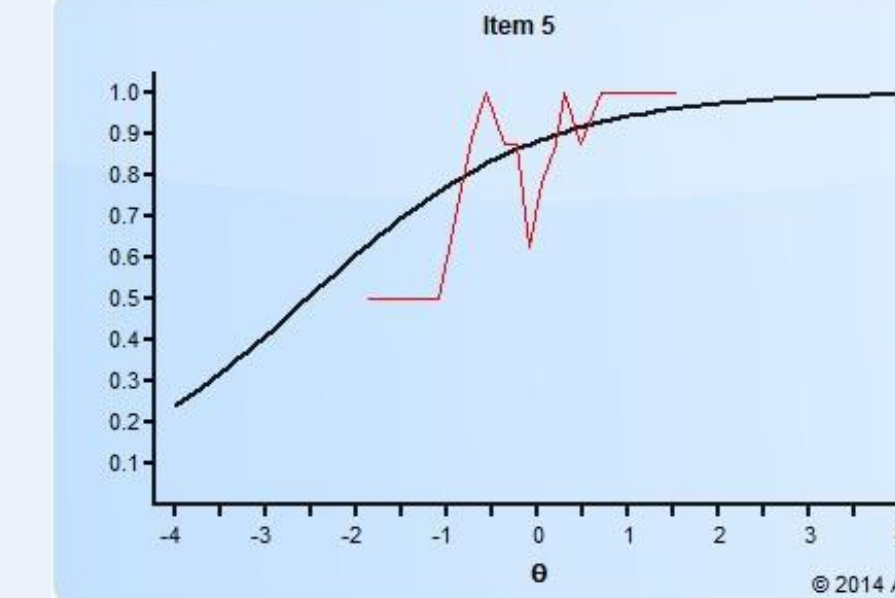
Note: Y-axis = proportion correct units; R Y-axis = number correct units

Figure 2b: Item Response Function



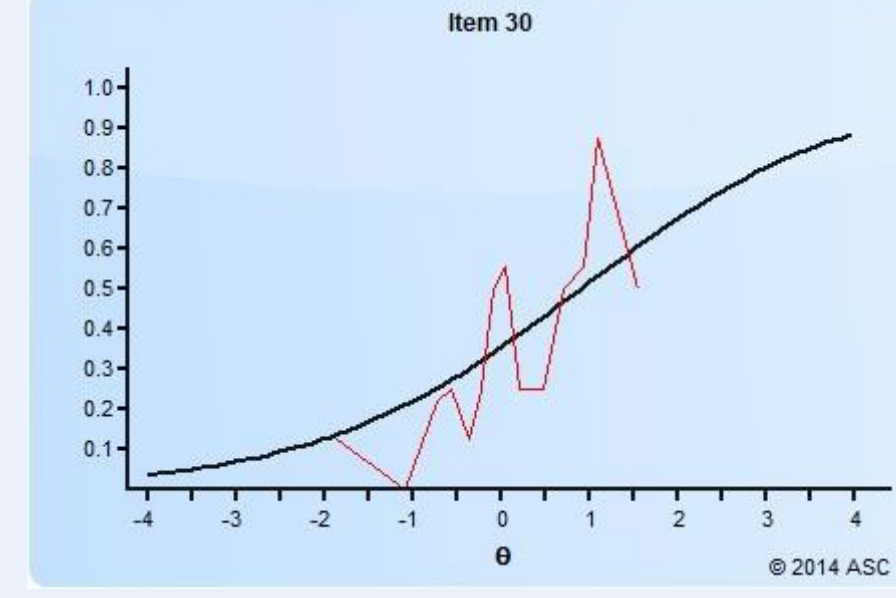
Strong Test Items

Figure 2c: Item Response Function



Note: Y-axis = proportion correct units; R Y-axis = number correct units

Figure 2d: Item Response Function



RESULTS:MATH

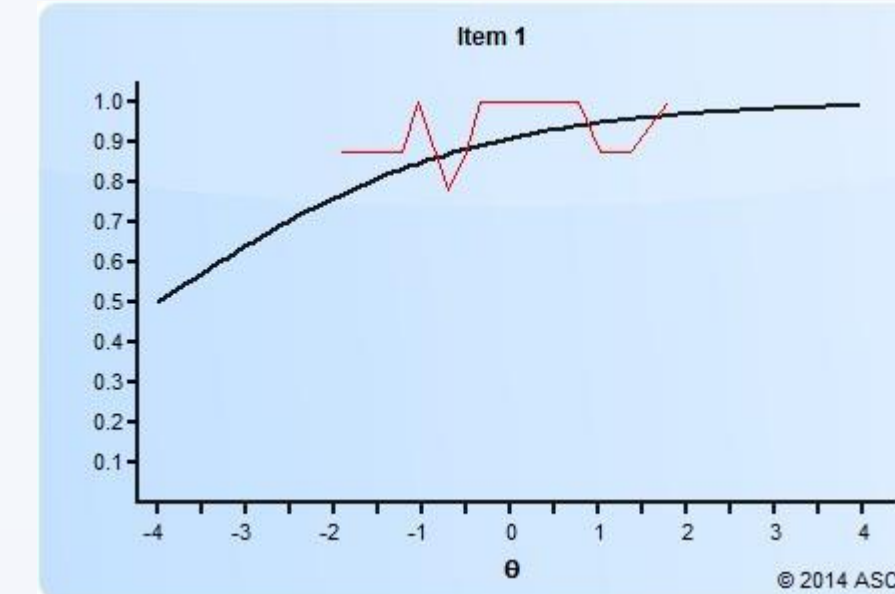
Table 3: Descriptive statistics for CERT Math test

Item	CTT		IRT 2PLM	
	P	R	a	b
1	0.943	0.088	0.568	-4.000
32	0.431	0.361	0.933	0.329
42	0.480	0.235	0.754	0.093
44	0.106	0.208	0.995	2.347

Full Test $M = 49.40$, $SD = 8.54$, Cronbach's $\alpha = .894$

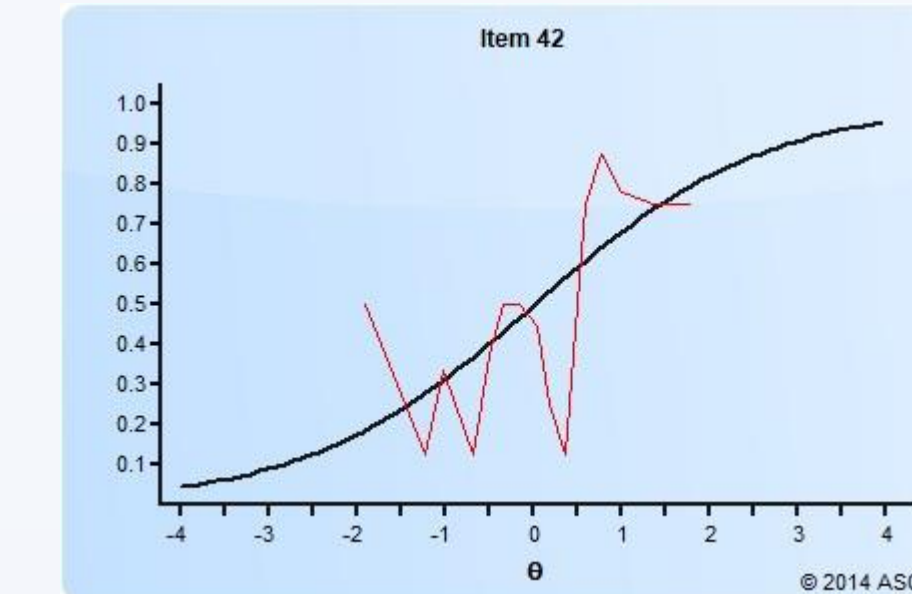
Weak Test Items

Figure 3a: Item Response Function



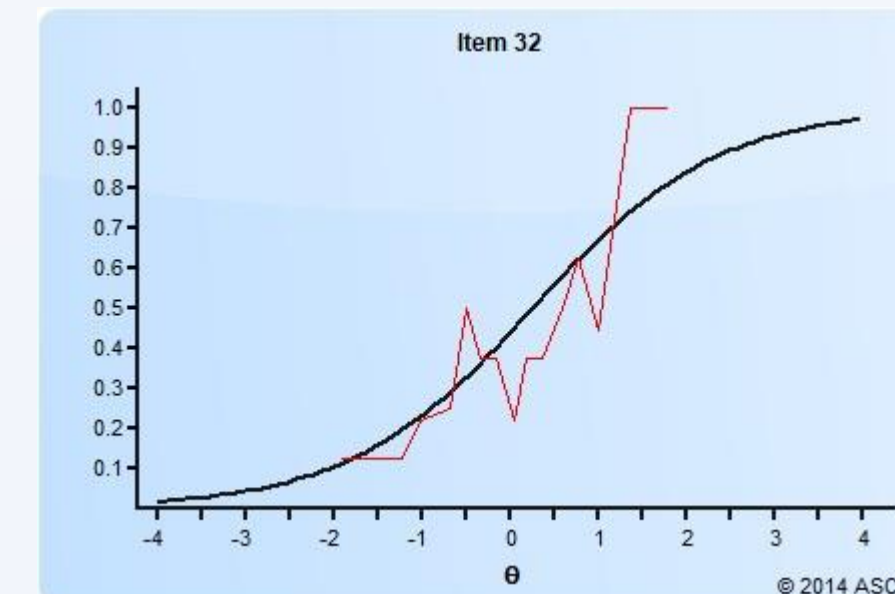
Note: Y-axis = proportion correct units; R Y-axis = number correct units

Figure 3b: Item Response Function



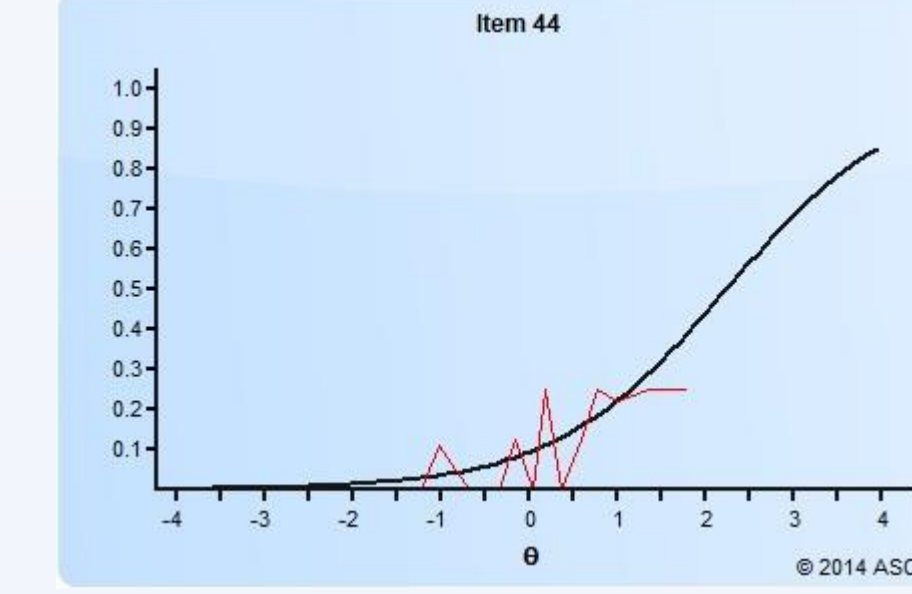
Strong Test Items

Figure 3c: Item Response Function



Note: Y-axis = proportion correct units; R Y-axis = number correct units

Figure 3d: Item Response Function



RESULTS: SCIENCE

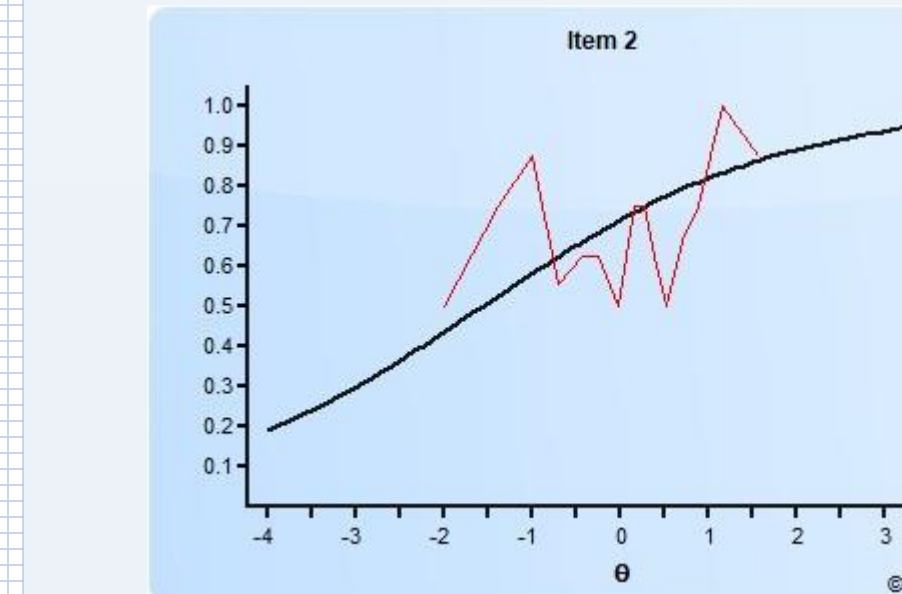
Table 4: Descriptive statistics for CERT Science test

Item	CTT		IRT 2PLM	
	P	R	a	b
1	0.902	0.326	1.139	-2.430
2	0.689	0.110	0.581	-1.515
9	0.463	0.085	0.565	0.254
36	0.333	0.328	0.857	0.917

Full Test $M = 25.98$, $SD = 6.83$, Cronbach's $\alpha = .848$

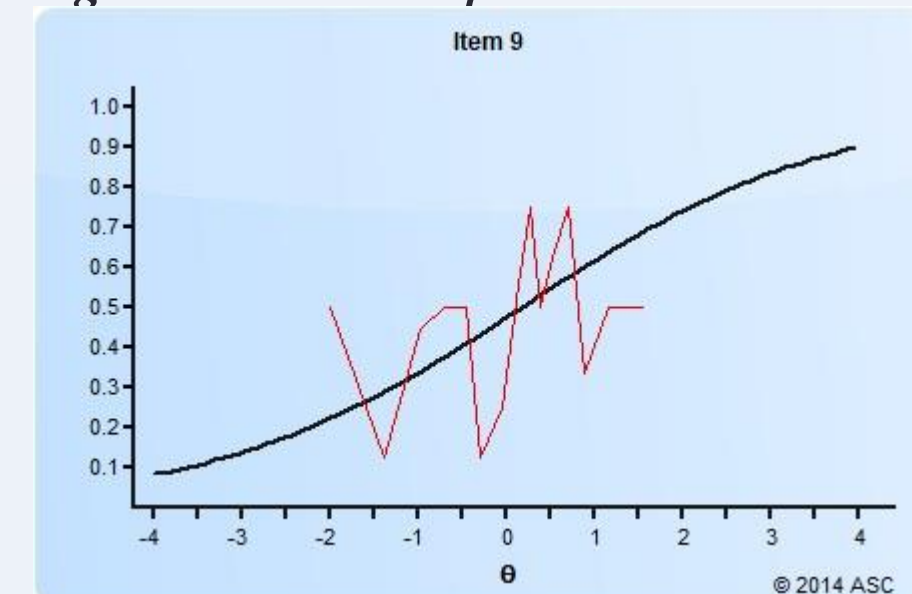
Weak Test Items

Figure 4a: Item Response Function



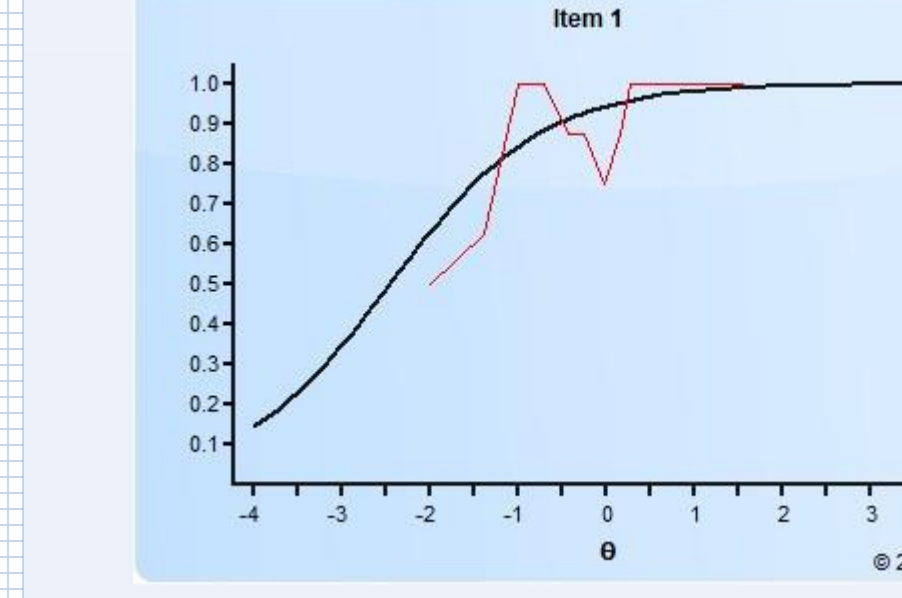
Note: Y-axis = proportion correct units; R Y-axis = number correct units

Figure 4b: Item Response Function



Strong Test Items

Figure 4c: Item Response Function



Note: Y-axis = proportion correct units; R Y-axis = number correct units

Figure 4d: Item Response Function

