Similarities and Differences Between Classical and Item Response Theory

This noon seminar by Ron D. Hays, Ph.D., is supported in part by the UCLA/DREW Project EXPORT, National Institutes of Health, National Center on Minority Health & Health Disparities, (P20-MD00148-01) and the UCLA Center for Health Improvement in Minority Elders / Resource Centers for Minority Aging Research, National Institutes of Health, National Institute of Aging, (AG-02-004).

Upcoming Conferences

October 17-19, *Applications of Item Response Theory to Health.* International Conference on Health Policy Research: Methodological Issues in Health Services and Outcomes Research, Chicago

Spring, 2004. NCI sponsored meeting, *Improving the Measurement of Cancer Outcomes through the Applications of Item Response Theory (IRT) Modeling: Exploration of Item Banks and Computer-Adaptive Assessment.* DC.

Reise and Henson (in press) Journal of Personality Assessment

"The critical question is not whether IRT models are superior to CTT methods. Of course they are, in the same way that a modern CD player provides superior sound when compared to a 1960s LP player…The real question is, does application of IRT result in sufficient improvement in the quality of … measurement to justify the added complexity?" Reise & Henson, J Personality Assessment, in press.

In the last 12 months, …

5. when you needed care from Dr. Shapiro for an illness or injury, how often did you get care as soon as you wanted?
7. how often did you get an appointment with Dr. Shapiro for regular or routine health care as soon as you wanted?
9. when you called Dr. Shapiro's office during regular office hours, how often did you get the help or advice you needed?
11. did the after hours care available to you from Dr. Shapiro meet your needs?

Never, Sometimes, Usually, Always (5, 7, 9)

No, Yes (11)

Basic Study Design

Patients were selected from physicians with at least 300 unique households with an encounter in the previous 12 months.

Patients were eligible if an adult member of 3 health plans or a large physician group in greater Cincinnati metro area and had at least one visit to one of the targeted physicians in the last 12 months.

3,804 surveys completed ($\overline{X}_{age}$ = 48; 59% female);

n = 351 with complete data on 4 access items

Evaluating Multi-item Scales

Scale Characteristics

- Reliability and unidimensionality
- Distribution of scores (level on attribute)

Item Characteristics

- Item difficulty
- Item-scale correlation ("discrimination")

Internal Consistency Reliability (alpha)

Source	df	SS	MS
Respondents	350	164.5	0.47
Items	3	11.88	3.96
Resp. x Items	1050	105	0.10
Total	1403		

$$\text{Alpha} = \frac{0.47 - 0.10}{0.47} = \boxed{0.78}$$

Standard Error of Measurement

$$SEM = S\,(1 - \text{reliability})^{1/2}$$

$$SEM = (.22)^{1/2} = 0.46$$

Person Score (level on attribute)

Average items together and compute z-score

Mean = 0, SD = 1, range: -2.07->0.83

(-2.07;-1.34; -0.62; 0.11; 0.83)

$$z_X = \frac{(X - \bar{X})}{SD_x}$$

Item difficulty (p = 0.84)

Proportion of people endorsing the item (p) can be expressed in z distribution form:

$$Z = \ln (1-p)/p)/1.7 = (\ln (1-p) - \ln (p))/1.7$$

$$= (\ln (.16) - \ln (.84))/1.7$$

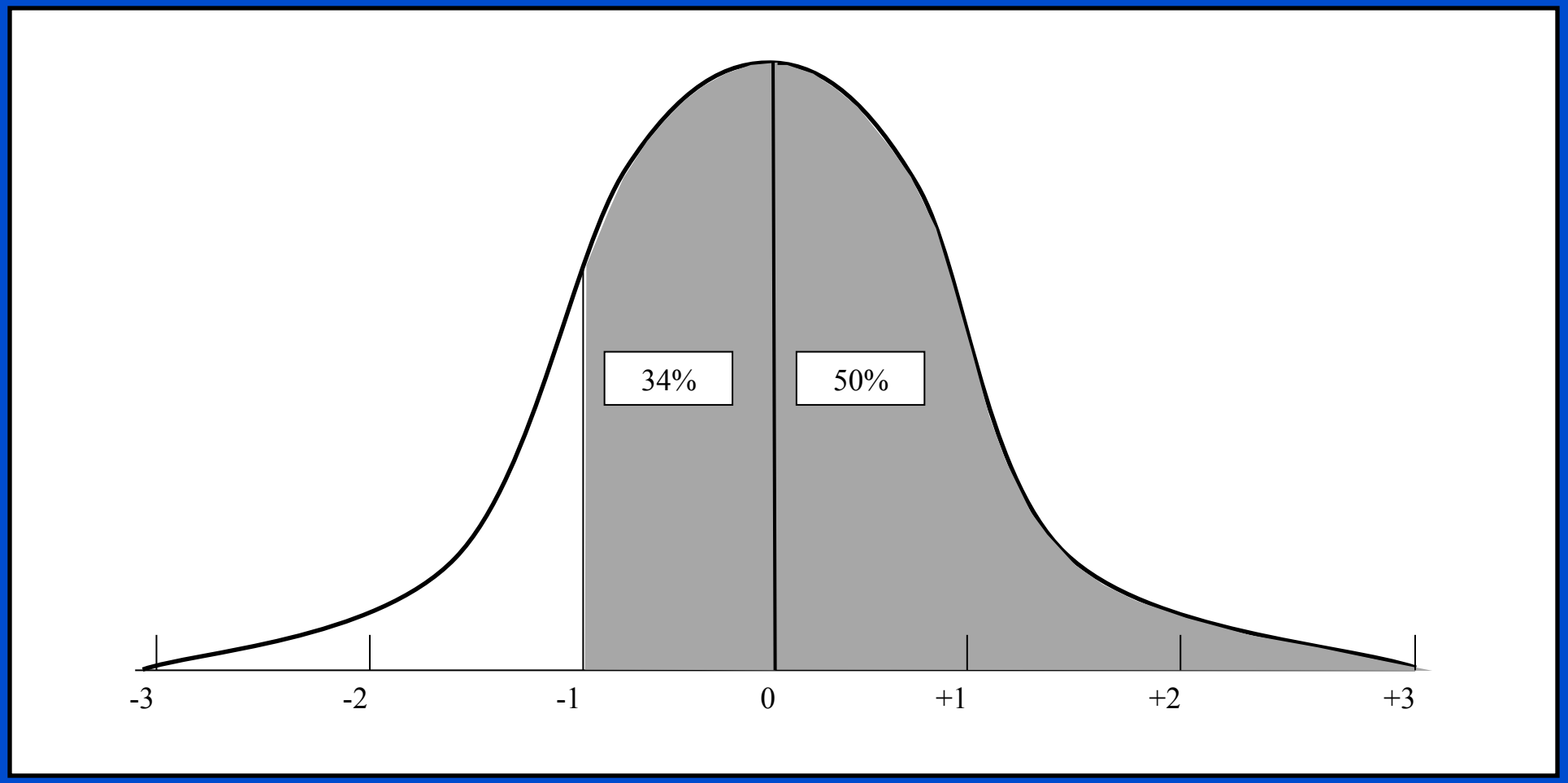
$$= (-1.83 + .17)/1.7$$

$$= -1.66/1.7$$

$$= \underline{-1.00}$$

(-2 -> 2 is typical range)

P-value transformation for an Item (*p*=.84)



Item difficulty (5): p = 0.68

How often did you get illness or injury care as soon as you wanted?

$z = \ln (1-p)/p)/1.7 = (\ln (.32) - \ln (.68))/1.7$

$= \underline{-0.43}$

Item difficulty (7): p = 0.61

How often did you get an appointment for regular or routine health care as soon as you wanted?

$z = \ln (1-p)/p)/1.7 = (\ln (.39) - \ln (.61))/1.7$

$= \underline{-0.26}$

Item difficulty (9): p = 0.71

How often when you called did you get the help or advice you needed?

$z = (\ln(1-p) - \ln(p))/1.7 = (\ln(.29) - \ln(.71))/1.7$

$= \underline{-0.52}$

Item difficulty (11): p = 0.86

Did the after hours care meet your needs?

$z = \ln (1-p)/p)/1.7 = (\ln (.14) - \ln (.86))/1.7$

$= \underline{-1.07}$

Item Difficulties

	p	z
Care for illness or injury	0.68	-0.43
Regular or routine care	0.61	-0.26<-
Office hour help/advice	0.71	-0.52
After hours care	0.86	-1.07<-

Item-Scale Correlations ("discrimination")

	Access scale
Care for illness or injury	0.69
Regular or routine care	0.61
Office hour help/advice	0.61
After hours care	0.46

Item-scale correlations are corrected for item overlap with the scale score.

Item-Scale Correlations

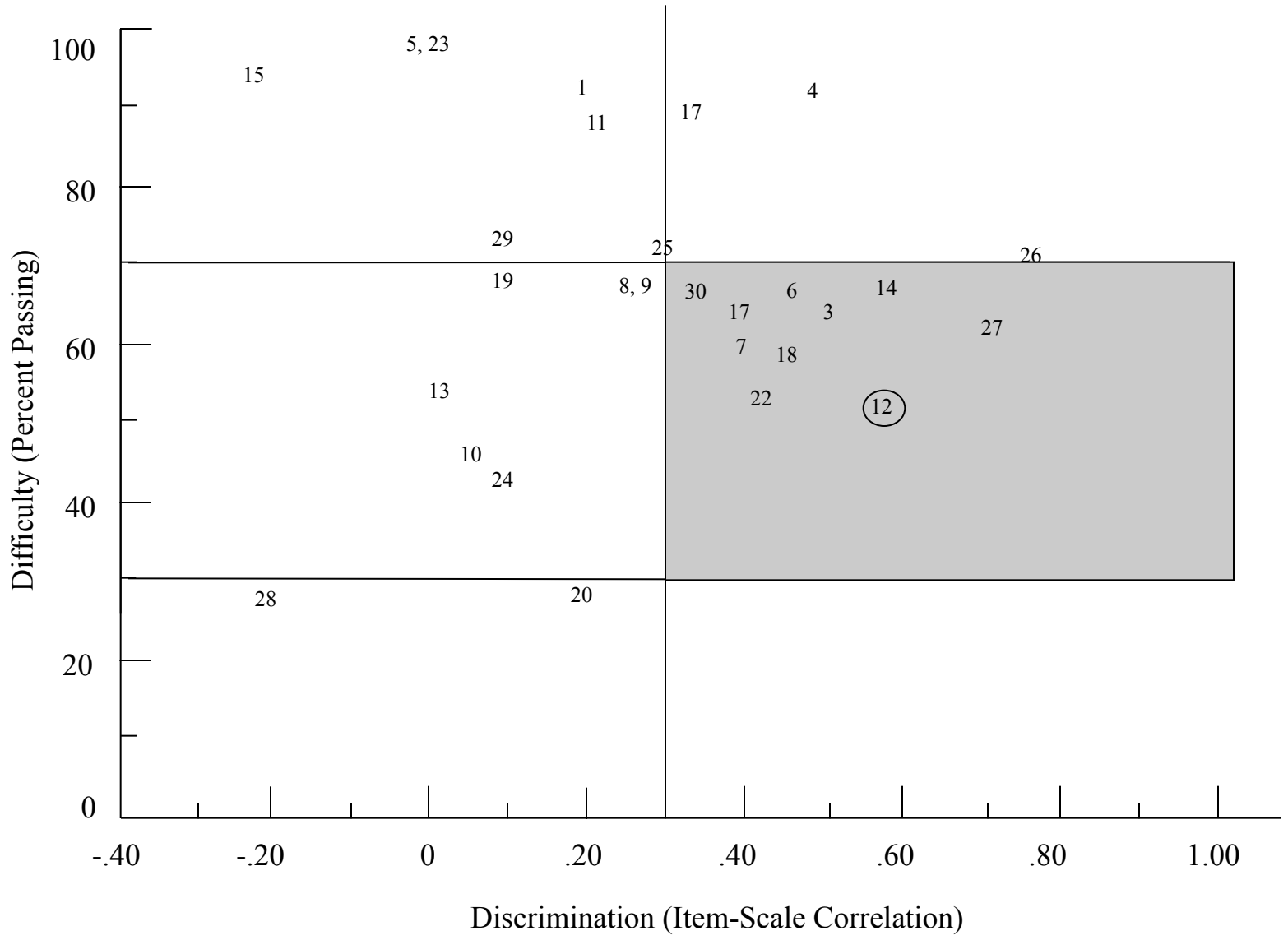
Item-scale correlation can be expressed in terms of z-statistic:

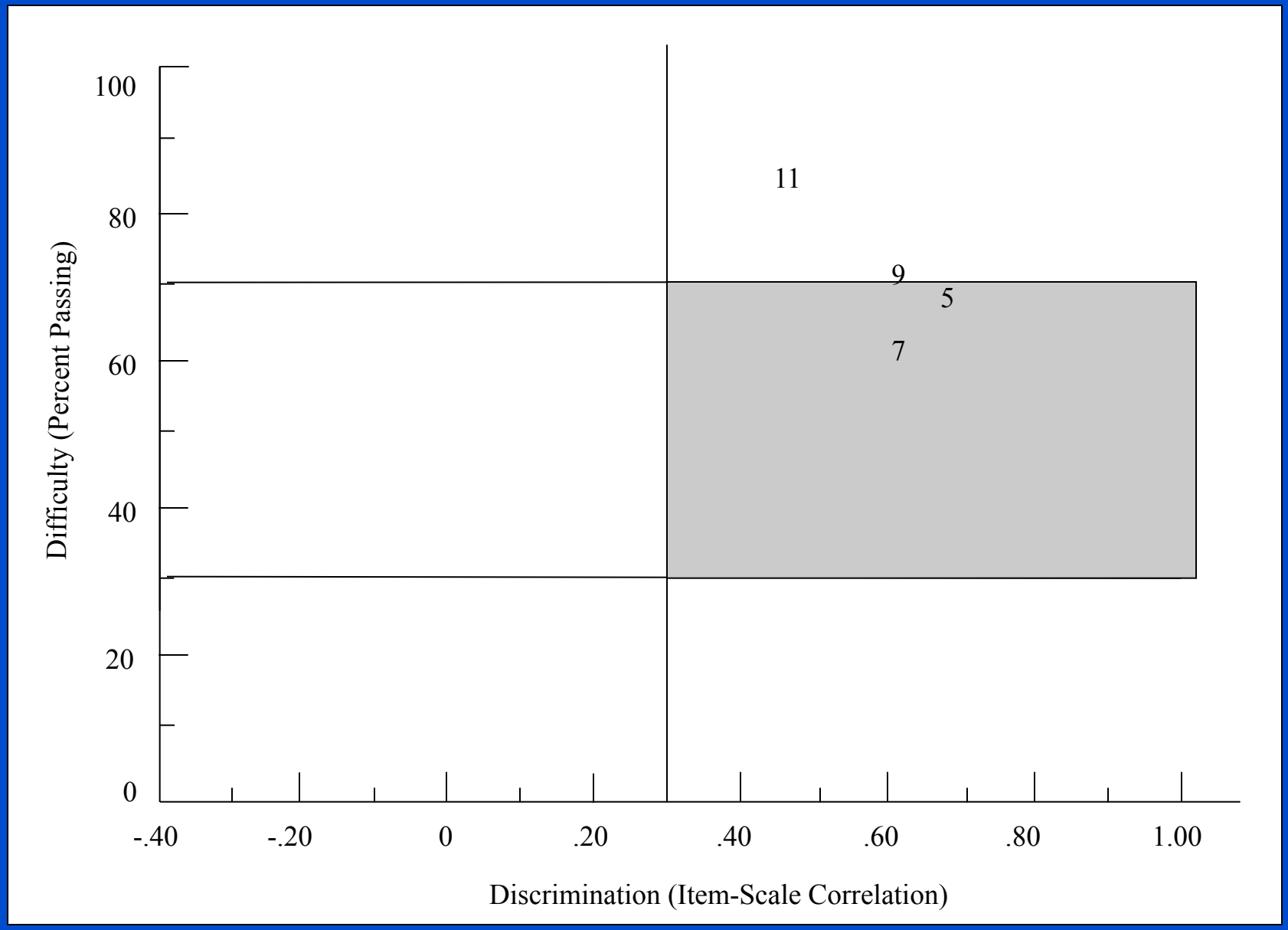
- $z = \frac{1}{2} [\ln (1 + r) - \ln (1-r)]$
- if r = 0.30, z = 0.31
- if r = 0.80, z = 1.10
- if r = 0.95, z = 1.83

Item-Scale Correlations

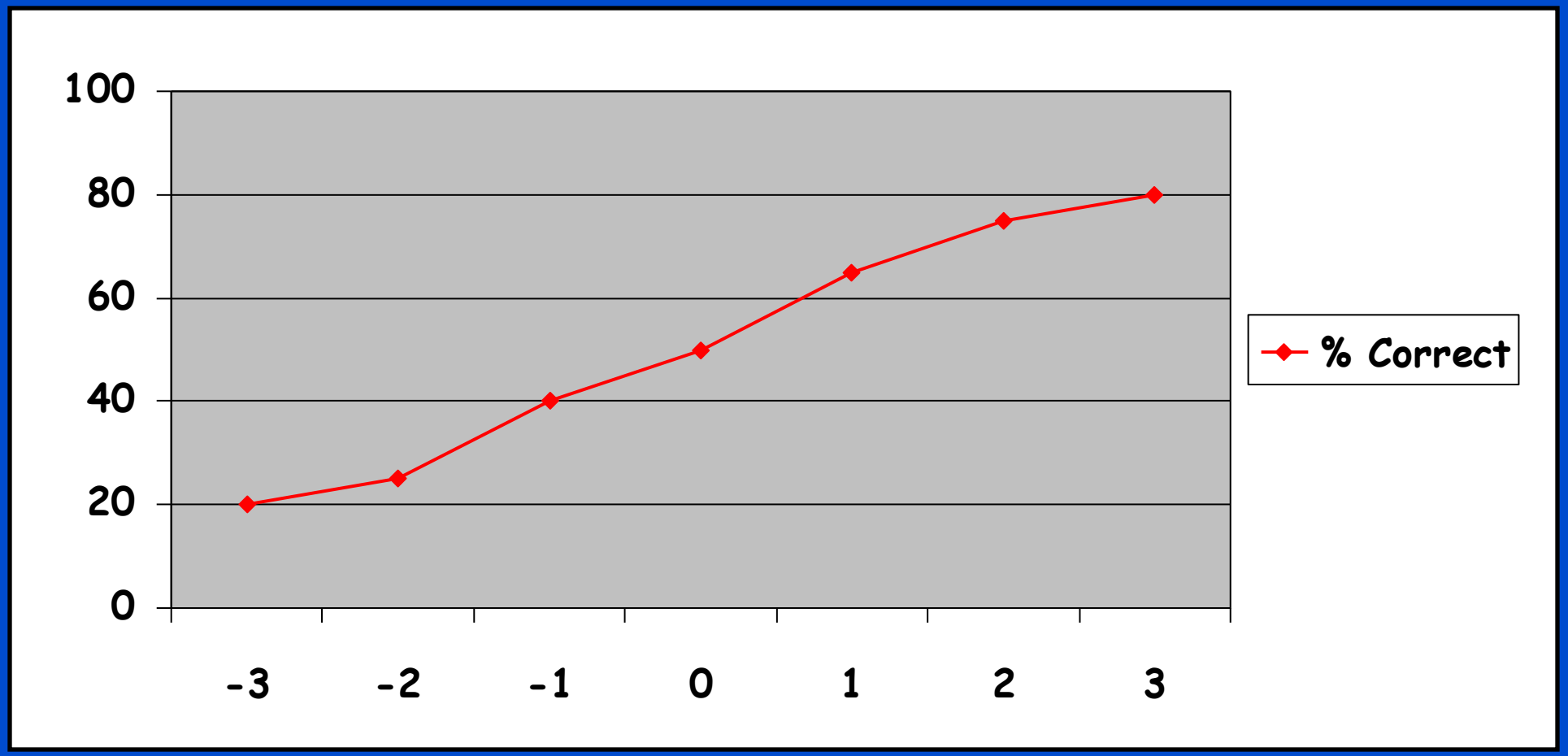
	Access scale	
	r	z
Care for illness or injury	0.69	0.85<-
Regular or routine care	0.61	0.71
Office hour help/advice	0.61	0.71
After hours care	0.46	0.50<-

Item-scale correlations are corrected for item overlap with the scale score.

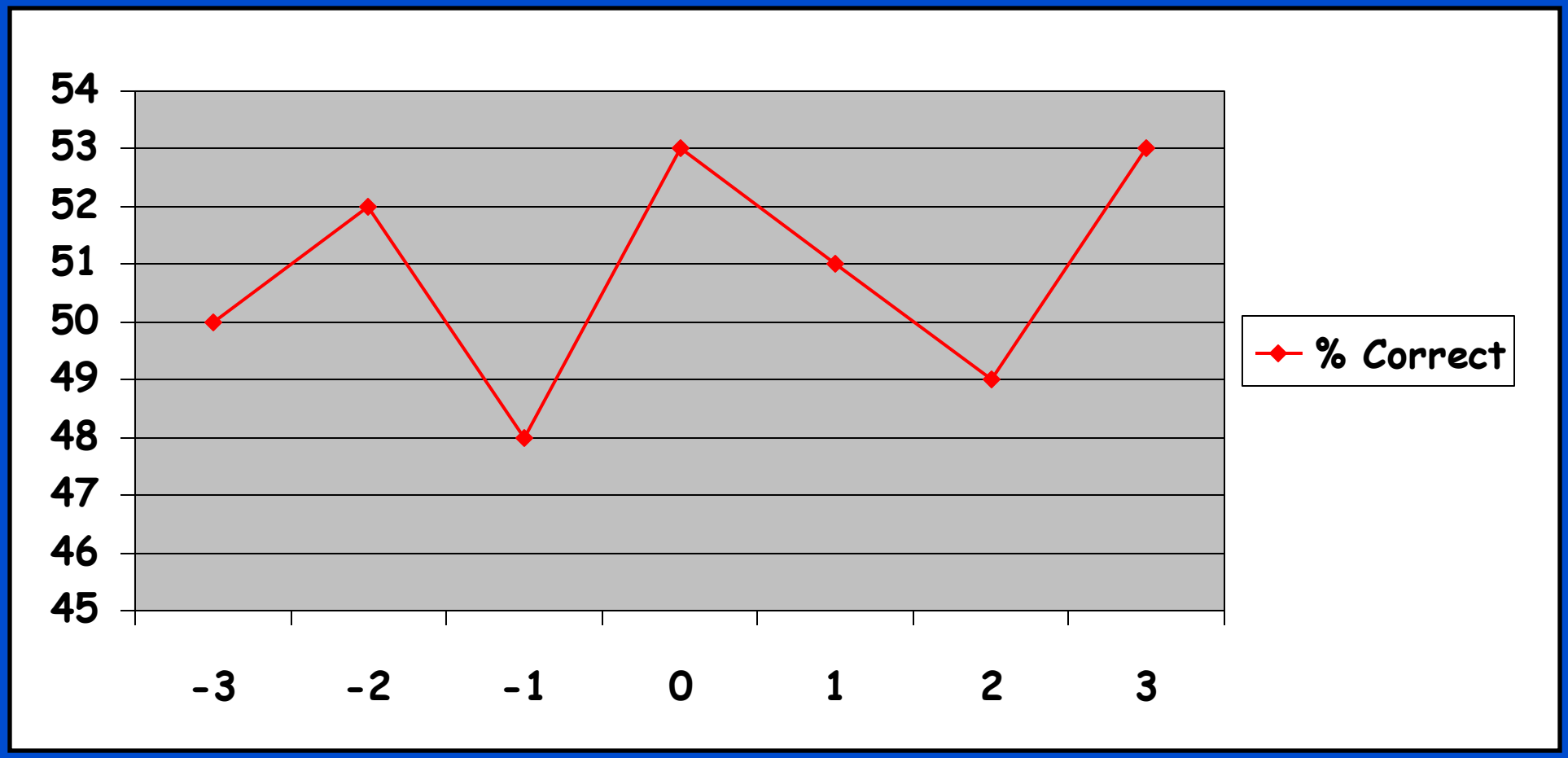

100
80
60
40
20
0
Difficulty (Percent Passing)
5, 23
15
1
11
17
4
29
25
26
19
8, 9
30
6
14
17
3
27
7
18
13
22
12
10
24
28
20
-.40
-.20
0
.20
.40
.60
.80
1.00
Discrimination (Item-Scale Correlation)

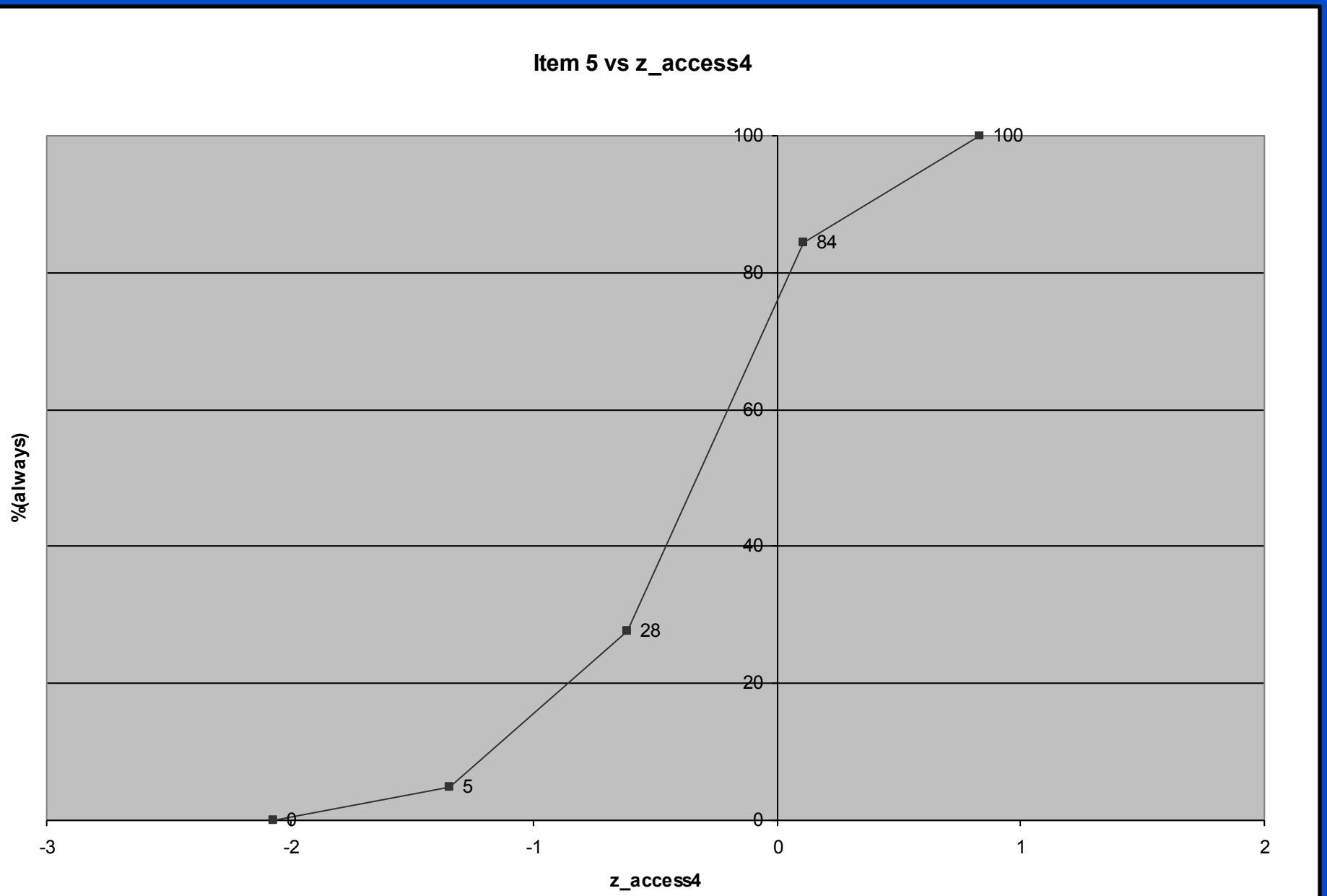

100
80
60
40
20
0
Difficulty (Percent Passing)
11
9
5
7
-.40
-.20
0
.20
.40
.60
.80
1.00
Discrimination (Item-Scale Correlation)

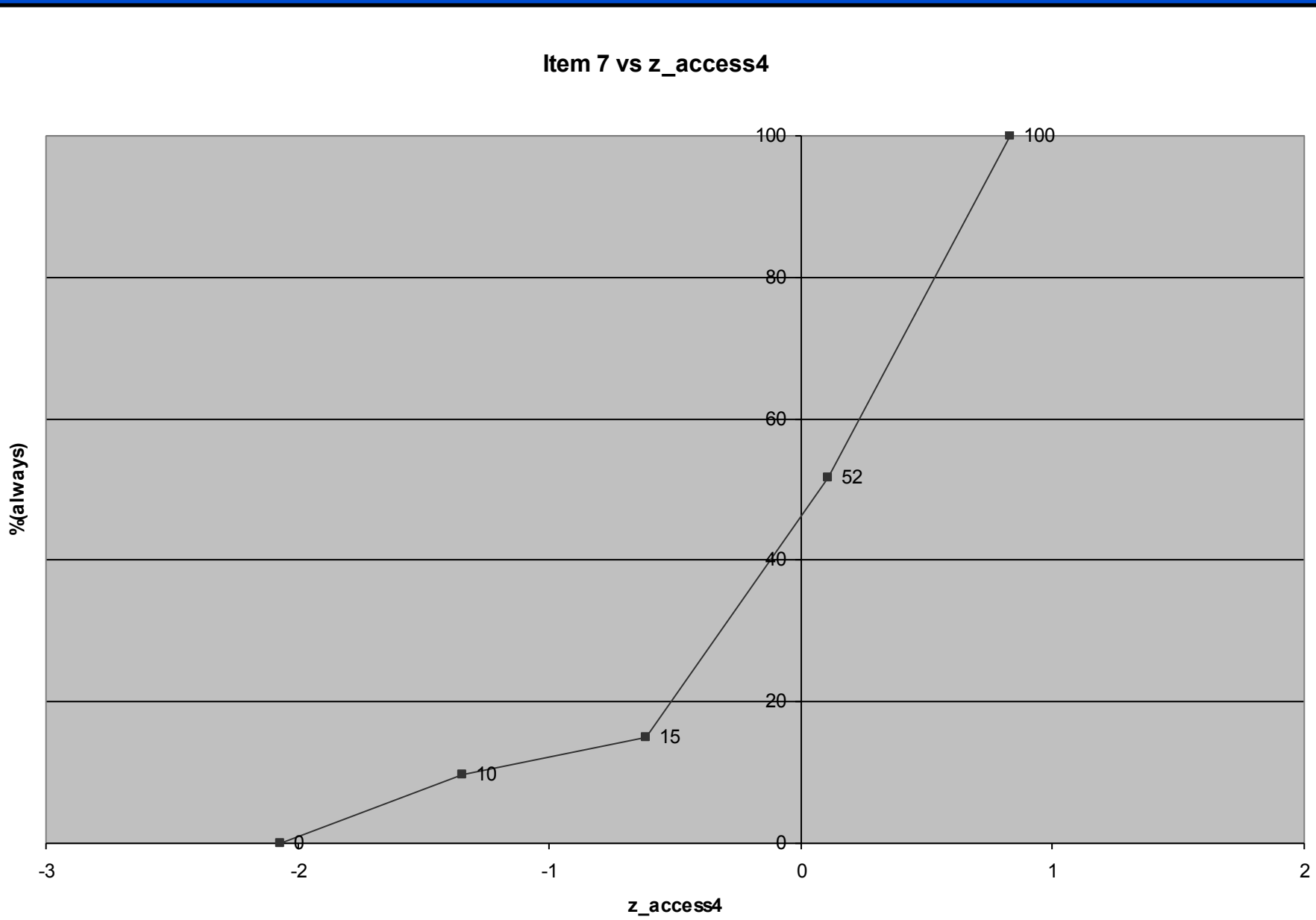
Item Characteristic Curve for a good item

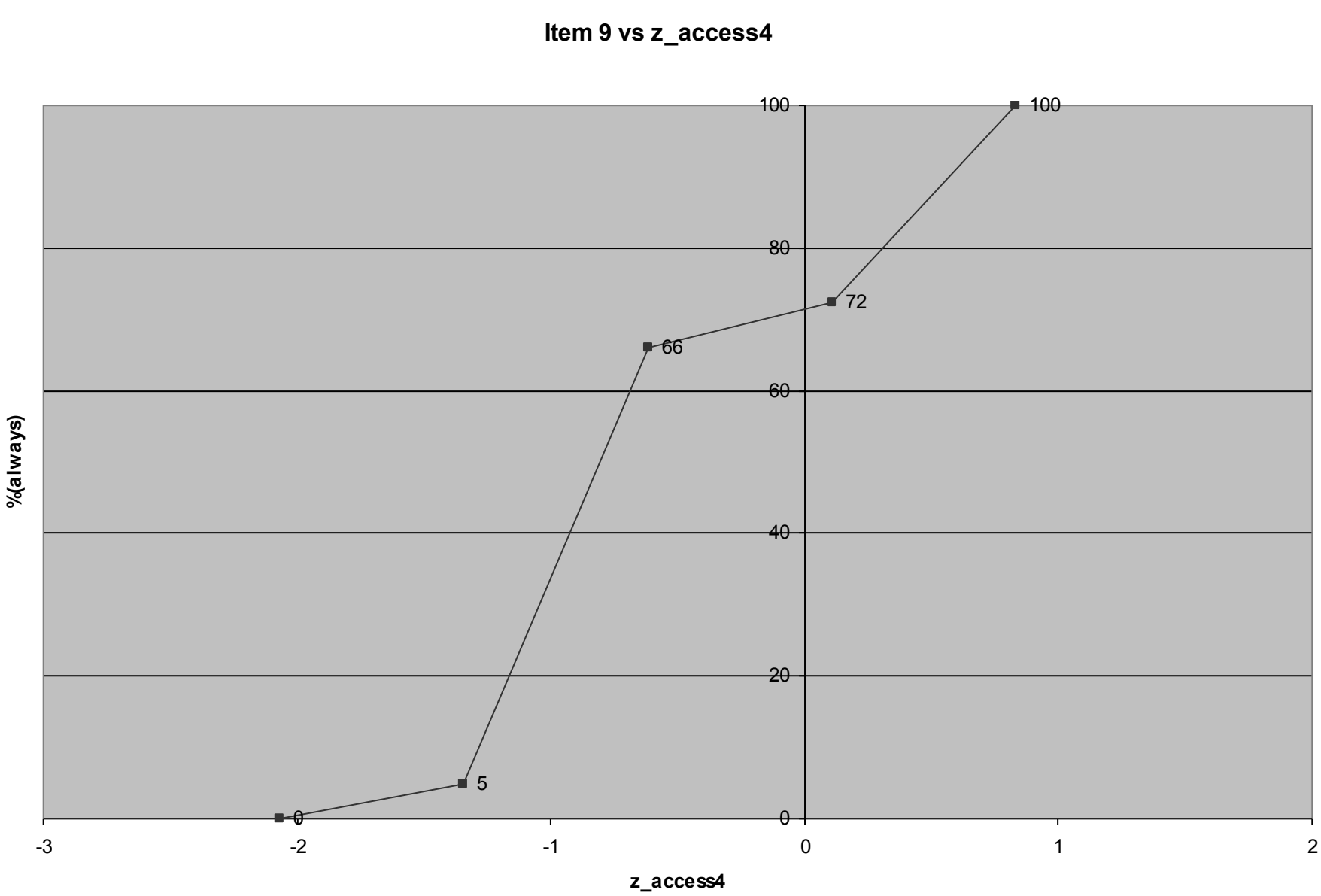


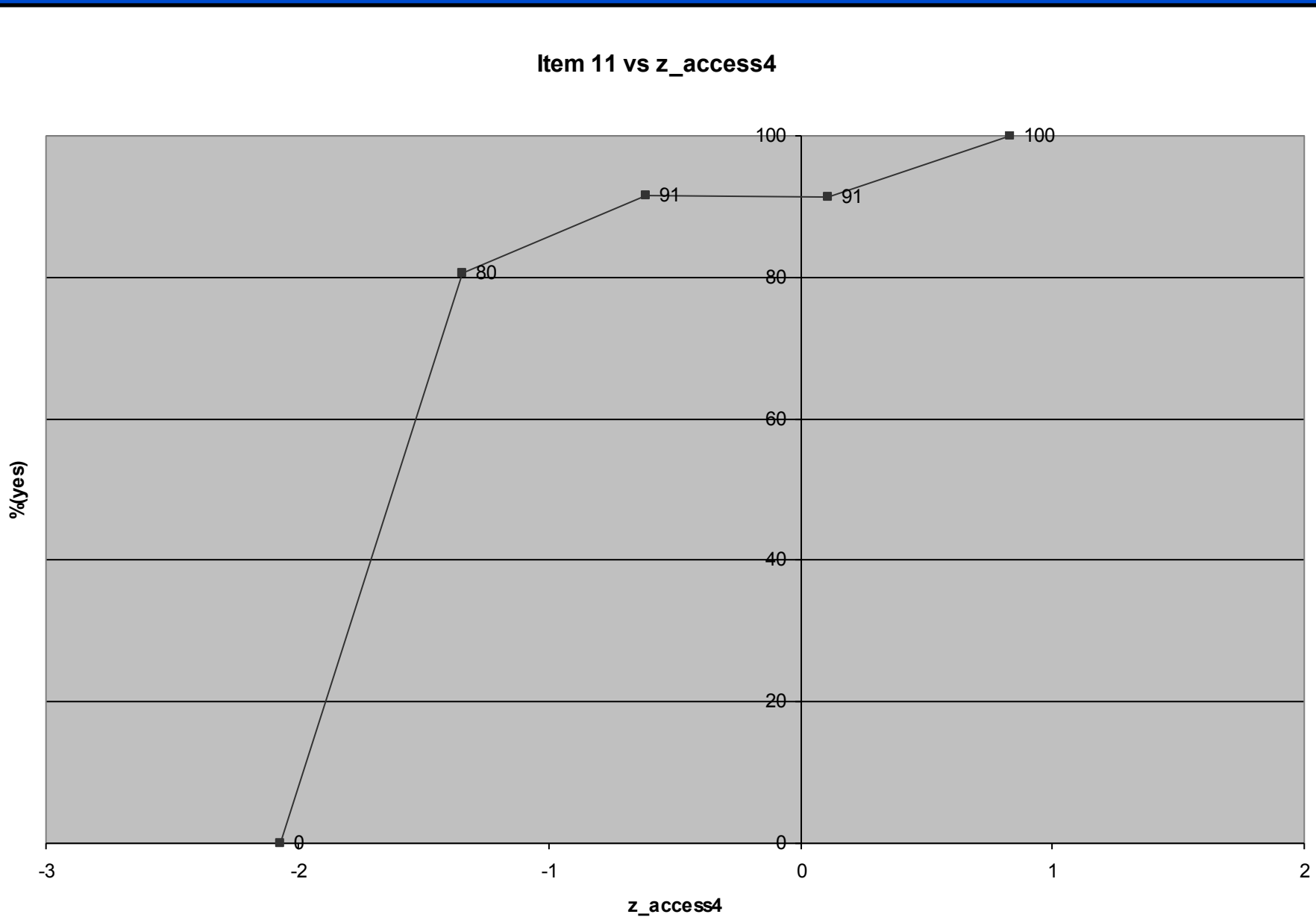
Item Characteristic Curve for poor item



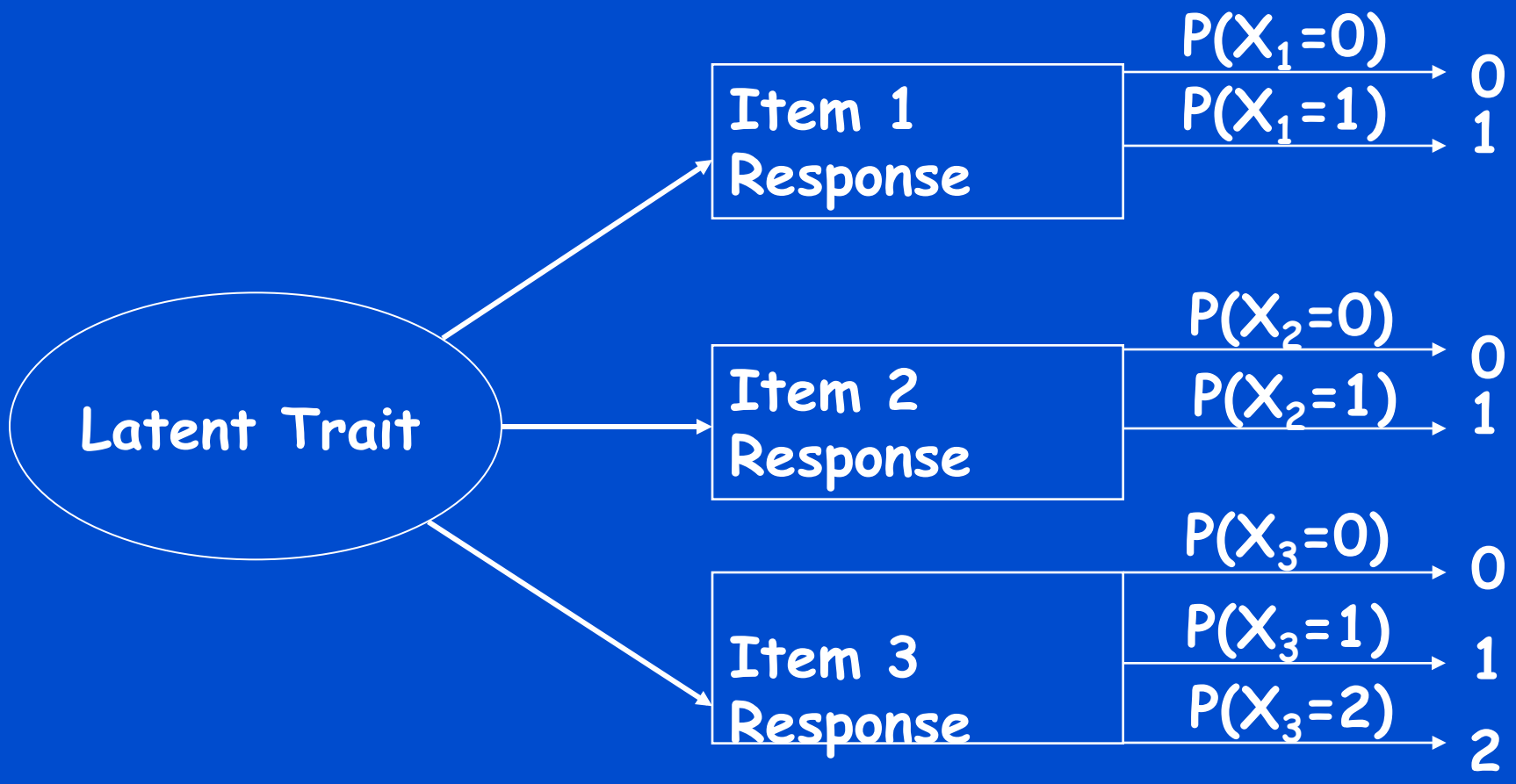

Item 5 vs z_access4
%(always)
100
80
60
40
20
0
-3
-2
-1
0
1
2
z_access4
0
5
28
84
100


Item 7 vs z_access4
%(always)
z_access4
100
80
60
40
20
0
-3
-2
-1
0
1
2
0
10
15
52
100

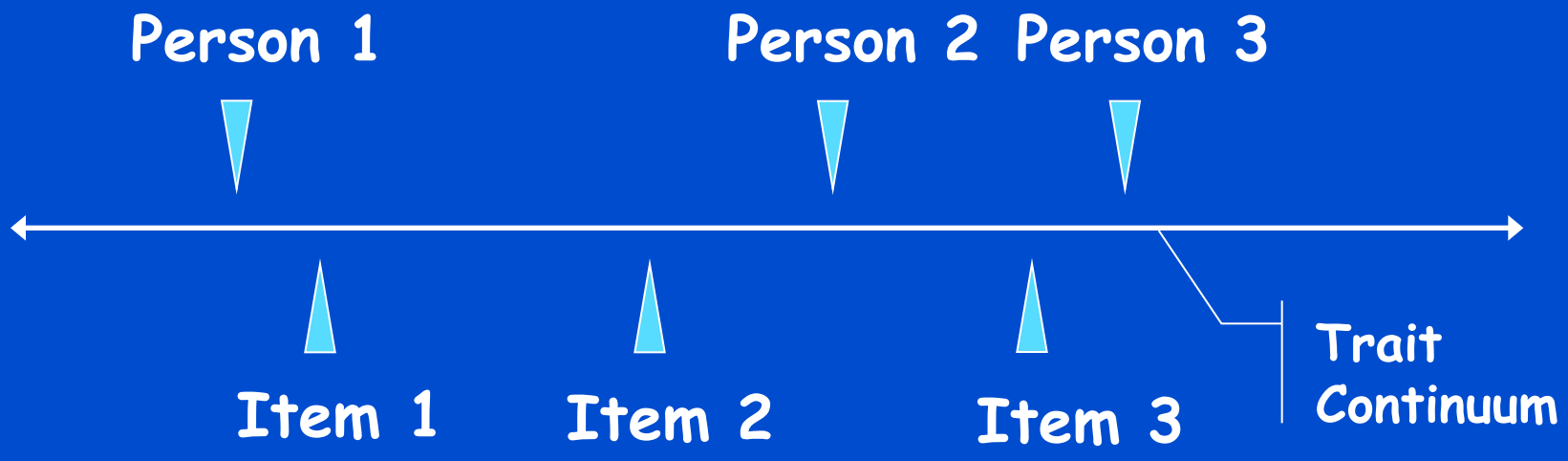

Item 9 vs z_access4
%(always)
z_access4
100
80
60
40
20
0
-3
-2
-1
0
1
2
0
5
66
72
100


Item 11 vs z_access4
%(yes)
z_access4
100
80
60
40
20
0
-3
-2
-1
0
1
2
0
80
91
91
100

Latent Trait and Item Responses



Item Responses and Trait Levels



IRT Versus CTT

- Item parameters (difficulty and discrimination) estimated using logistic models instead of proportions and item-scale correlations
- Variety of IRT models
  - 1, 2, and 3 parameter models
  - Dichotomous and polytomous
  - Graded response, partial credit, rating scale

2-Parameter Logistic Model

$$P_i(\Theta) = \frac{e^{1.7\,a_i\,(\Theta - b_i)}}{1 + e^{1.7\,a_i\,(\Theta - b_i)}}$$

$P_i(\Theta)$ Probability that a randomly selected respondent with ability Θ (trait level) answers "yes."

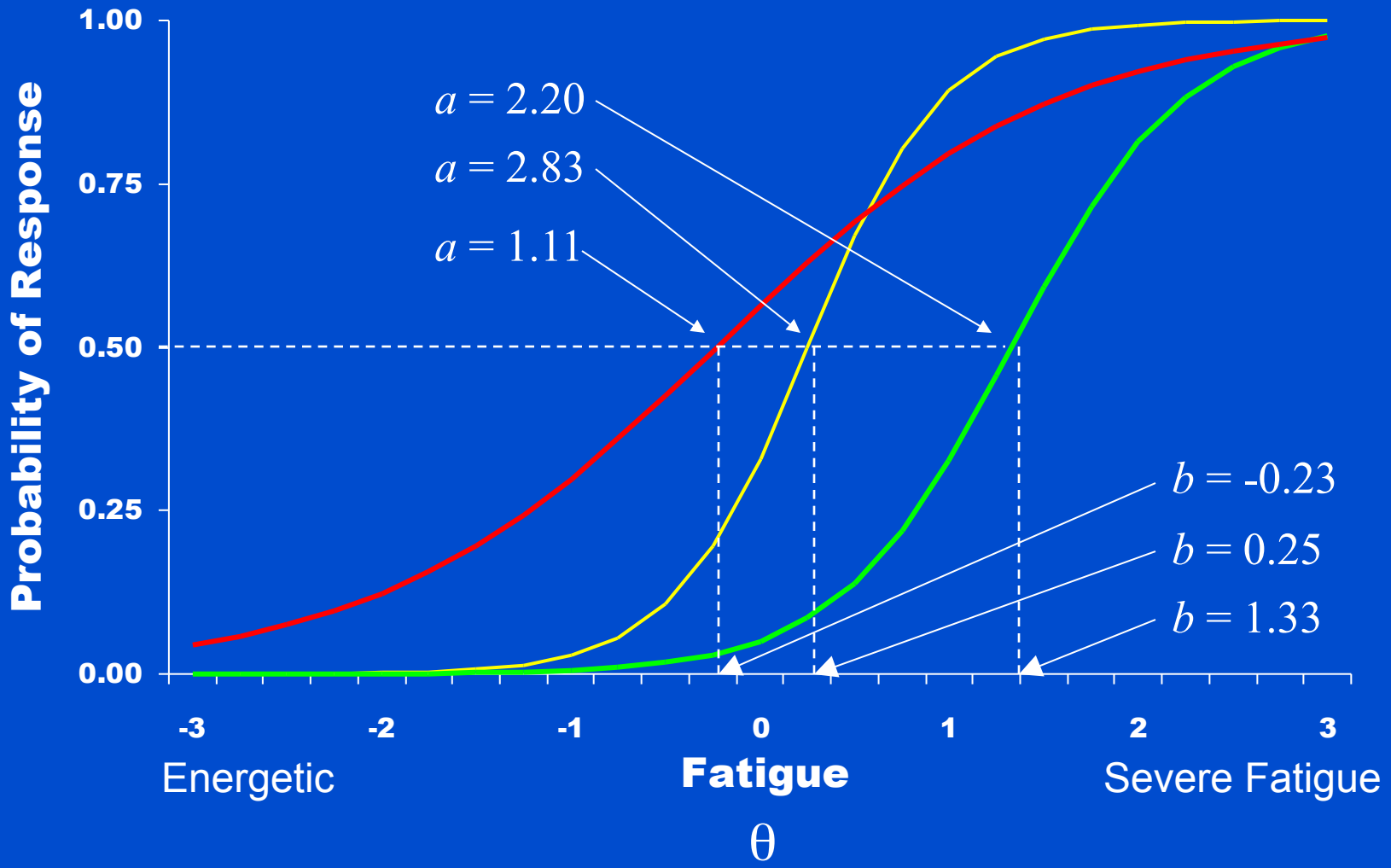
b_i Item i difficulty.

a_i Item i slope.

2-Parameter Logistic IRT Model

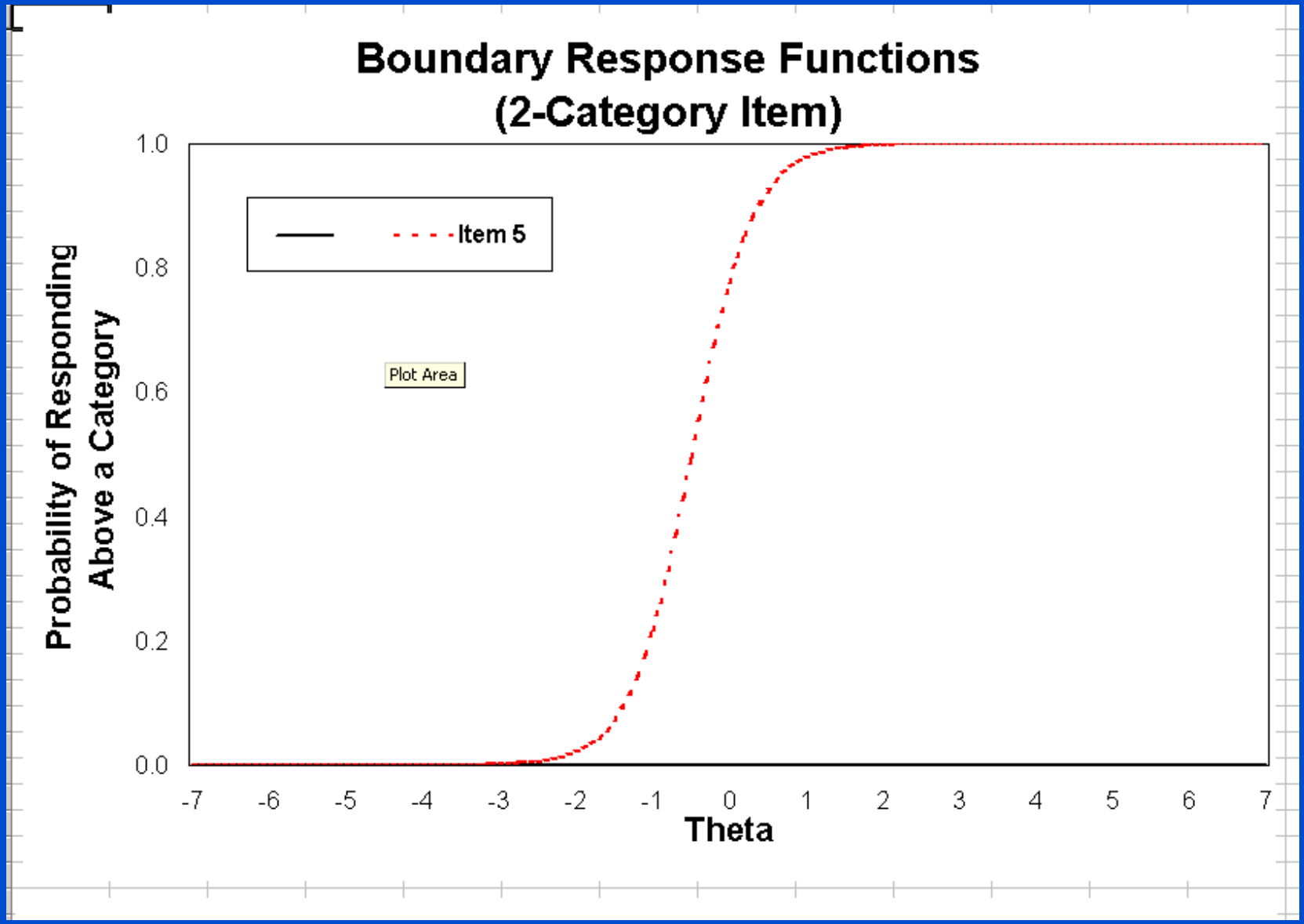
$$P(X_i = 1|\theta) = \frac{1}{1 + e^{-1.7a_i(\theta - b_i)}}$$

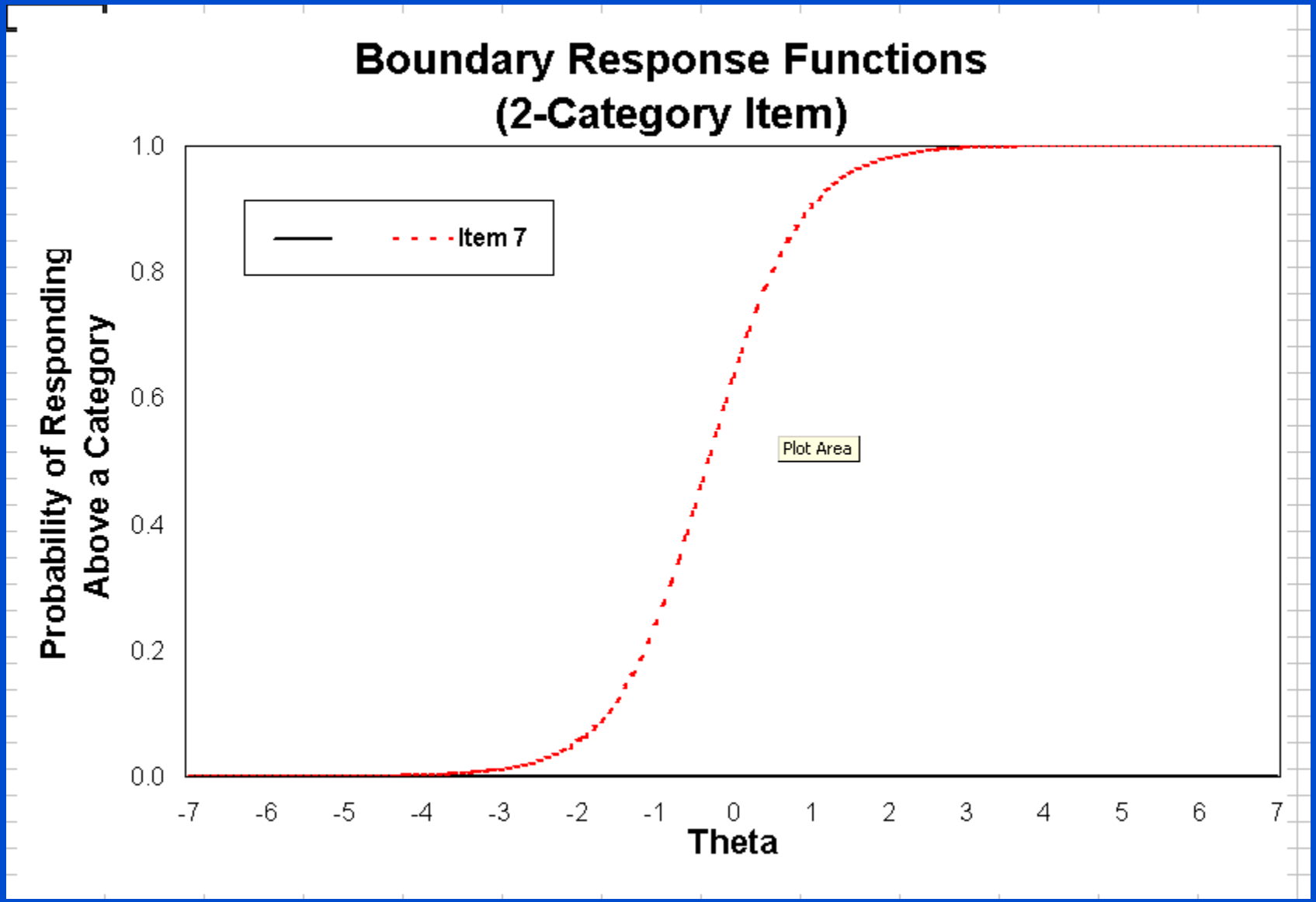
$a_i(\theta - b_i)$

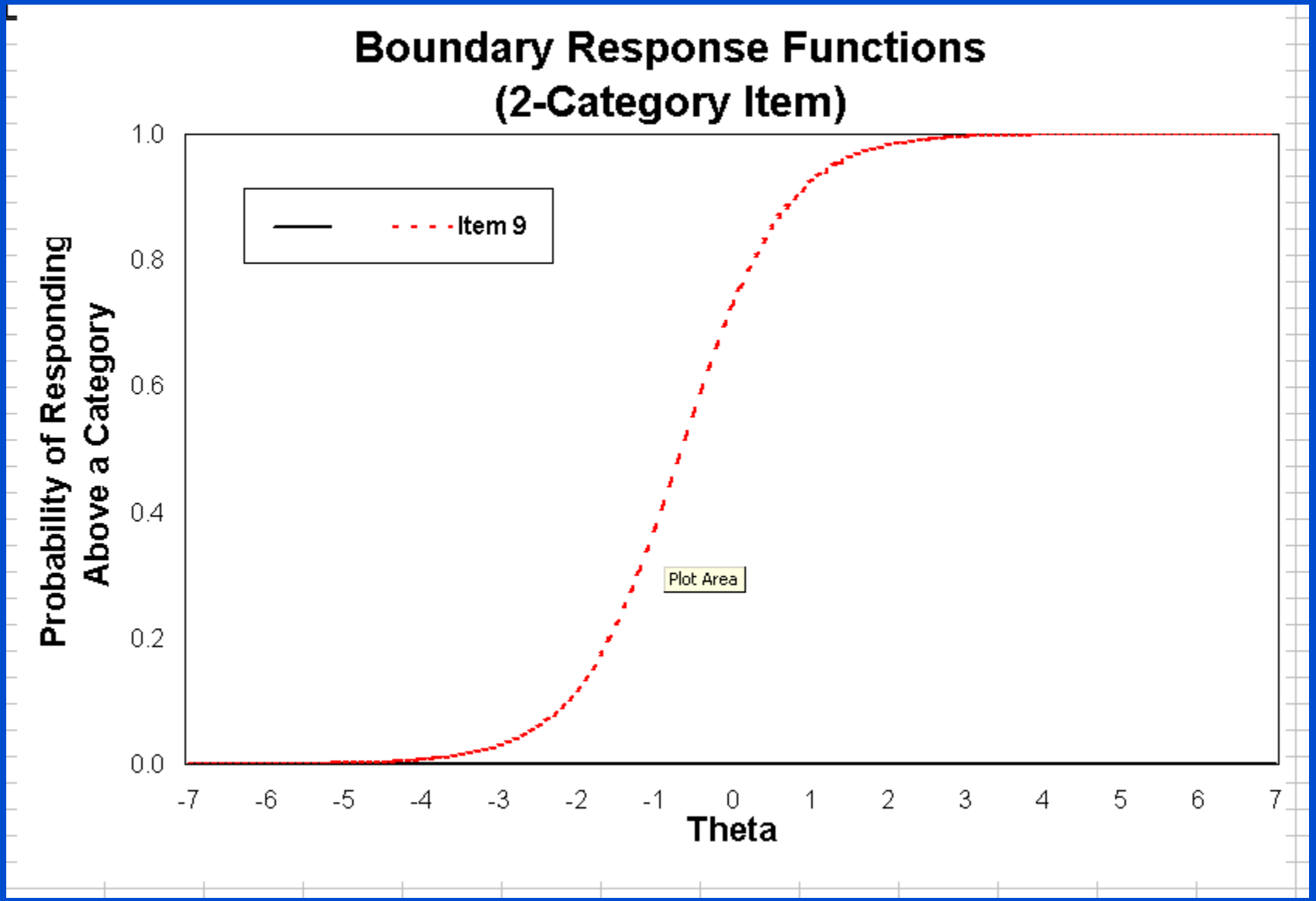


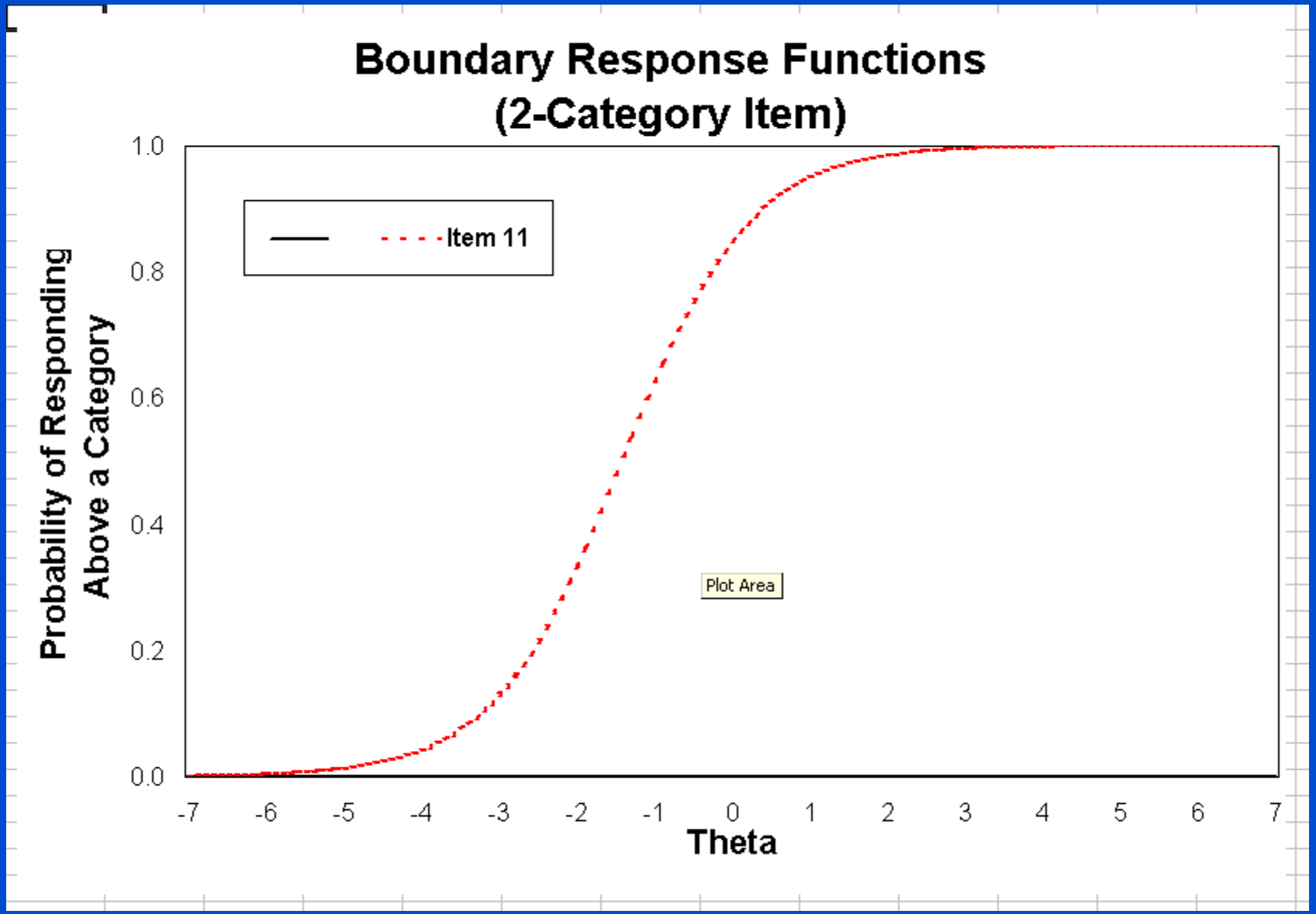
Difficulty and Discrimination Parameters

	Difficulty		Discrimination			
	CTT	IRT	CTT-IRT		CTT-IRT	
Care for illness or injury	-0.43	–0.49	0.85	0.93	2.00	2.53
Regular or routine care	-0.26	–0.32	0.71	0.86	1.32	1.69
Office hour help/advice	-0.52	–0.66	0.71	0.83	1.35	1.50
After hours care	-1.07	–1.43	0.50	0.77	1.05	1.21


Boundary Response Functions
(2-Category Item)
Item 5
Plot Area
Probability of Responding
Above a Category
1.0
0.8
0.6
0.4
0.2
0.0
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
Theta


Boundary Response Functions
(2-Category Item)
Item 7
Plot Area
Probability of Responding
Above a Category
1.0
0.8
0.6
0.4
0.2
0.0
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
Theta


Boundary Response Functions
(2-Category Item)
Item 9
Probability of Responding
Above a Category
1.0
0.8
0.6
0.4
0.2
0.0
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
Theta
Plot Area


Boundary Response Functions
(2-Category Item)
Item 11
Probability of Responding
Above a Category
1.0
0.8
0.6
0.4
0.2
0.0
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
Theta
Plot Area

IRT Versus CTT

- Reliability (information) conditional on underlying ability or attribute vs.
- Reliability estimated overall

Information Conditional on Trait Level

- Item information proportional to inverse of standard error of measurement:

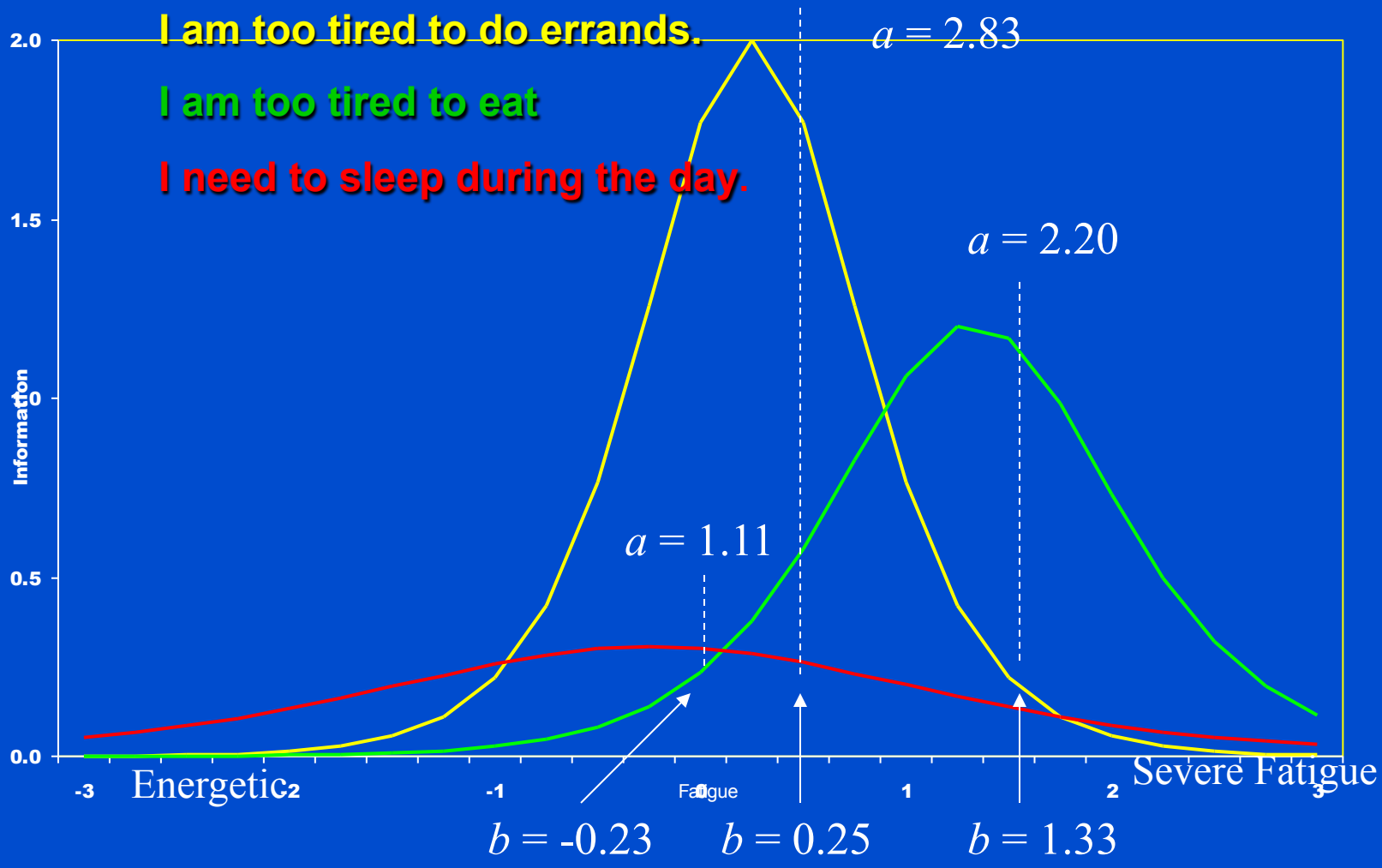
$$SEM(\Theta) = \frac{1}{\sqrt{I(\Theta)}}$$

- Scale information is the sum over item information:

$$I(\Theta) = \sum_{i=1}^{n} I_i(\Theta)$$

Item Information Curves

(The range of the latent construct over which an item is most useful for distinguishing among respondents)



IRT Versus CTT

- Item and person parameters incorporated into the same model.
- Marginal maximum likelihood estimation (MML) used to calibrate item parameters
- Level of attribute estimated by ML or Bayes methods rather than item sums

Scoring All Response Patterns Using Sum Score and Different IRT Models

More Fatigued

#	Item Response Pattern 0 = false, 1 = true	Summed Score	1 PL IRT / Rasch Model M-L Estimate	2 PL IRT Model M-L Estimate
1	0 0 0 0	0	-0.84	-0.82
2	1 0 0 0	1	-0.22	-0.27
3	0 1 0 0	1	-0.22	-0.21
4	0 0 1 0	1	-0.22	-0.19
5	0 0 0 1	1	-0.22	-0.01
6	1 1 0 0	2	0.22	0.14
7	1 0 1 0	2	0.22	0.15
8	0 1 1 0	2	0.22	0.19
9	1 0 0 1	2	0.22	0.31
10	0 1 0 1	2	0.22	0.36
11	0 0 1 1	2	0.22	0.37
12	1 1 1 0	3	0.71	0.52
13	1 1 0 1	3	0.71	0.72
14	1 0 1 1	3	0.71	0.74
15	0 1 1 1	3	0.71	0.80
16	1 1 1 1	4	1.36	1.35

IRT Strengths

CAT

Linking of scale

DIFF

IRT Versus CTT

- Interest in person fit as well as item fit
- Z_L has expected value of zero, with variance of one (if person responds according to the estimated IRT model). Large negative Z_L values (>= 2.0) indicate misfit.
- **Limited a lot** in feeding, getting around, preparing meals, shopping, and climbing one flight of stairs; but **limited a little** in vigorous activities, walking one block, and walking more than a mile.

- $Z_L = -9.56$

Worthwhile URLs

http://appliedresearch.cancer.gov/areas/cognitive/immt.pdf

http://work.psych.uiuc.edu/irt/

http://www.ssicentral.com/home.htm

Suggested Reading List

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. New Jersey: Erlbaum.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston: Kluwer-Nijhoff.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st Century. Medical Care, 38, II-28-42.

Thissen, D., & Wainer, H. (eds.). Test scoring. New Jersey, Erlbaum.