DOCUMENT RESUME

ED 415 281		TM 028 016
AUTHOR	Cantrell, Catherine E.	
TITLE	Item Response Theory: Understanding the One-Parameter Rasch Model.	
PUB DATE	1997-01-23	
NOTE	42p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 23, 1997).	
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Evaluative (142) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC02 Plus Postage.	
DESCRIPTORS	*Ability; *Difficulty Level; *Estimation (Mathematics); *Item Response Theory; *Mathematical Models; Prediction; Sampling; Tables (Data)	
IDENTIFIERS	Item Characteristic Function; Item Discrimination (Tests); *One Parameter Model; *Rasch Model	

ABSTRACT

This paper discusses the limitations of Classical Test Theory, the purpose of Item Response Theory/Latent Trait Measurement models, and the step-by-step calculations in the Rasch measurement model. The paper explains how Item Response Theory (IRT) transforms person abilities and item difficulties into the same metric for test-independent and sample-independent comparisons. IRT is based on two postulates. First, the performance of an examinee on a test item can be predicted by a set of factors called traits. Second, the relationship between examinee test performance and the set of traits underlying the performance can be defined by an item characteristic curve. There are three models in IRT. The three-parameter model is made up of item discrimination, item difficulty, and guessing parameters. In the one-parameter Rasch model, the guessing and item discrimination parameters are considered negligible. This model is used to analyze differences in test scores that initially are not linear. In this study, a regression analysis was performed to find a correlation between Rasch calibrations and classical measurement calibrations and to plot a scatterplot of the two measures with their regression line. These results challenge the idea that Rasch latent trait measurement is superior to classical measurement because its estimates are item-free and sample-free. High correlations between the two measures indicate that Rasch calibrations are not truly item- or sample-free or that the classical model calibrations are equally item- or sample-free. (Contains 12 tables, 6 figures, and 5 references.) (SLD)

Running head: ITEM RESPONSE THEORY

Item Response Theory:
Understanding the One-Parameter Rasch Model

Catherine E. Cantrell

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 23, 1997.

Abstract

The present paper discusses the limitations of Classical Test Theory, the purpose of Item Response Theory/Latent Trait Measurement models, and the step-by-step calculations in the Rasch measurement model. The present paper explains how IRT transforms person abilities and item difficulties into the same metric for test-independent and sample-independent comparisons.

Item Response Theory: Understanding the One-Parameter Rasch Model

Item Response Theory (IRT) or Latent Trait Theory came about due to the limitations of classical measurement models. Classical measurement defines person ability, also known as the true score, as the expected value of performance on a test. The problem with the classical definition is that ability estimate depends on the difficulty of the items chosen for the test. In other words, the ability estimate or score is test dependent. Likewise, item difficulty--defined by the classical theory as the proportion of examinees answering the item correctly-- depends on the ability of the particular people taking the test. In other words, item difficulties are group dependent (Hambleton & Swamination, 1985). Therefore, items and examinees on different tests are measured on different scales. Because classical theory item difficulties and person abilities are on different scales, it is inappropriate to compare them (Wright & Stone, 1979). Item Response Theory, on the other hand, transforms item difficulty and person ability estimates into statistics on a single comparable scale that are also respectively "person-free" and "item-free." "Person-free" means that the item difficulty calibrations are theoretically independent of the persons generating the calibrations; "item-free" means that the person ability estimates are theoretically independent of the items used on the measurement.

IRT is based on two postulates. First, the performance of an examinee on a test item can be predicted by a set of factors called traits, latent traits, or abilities. Second, the relationship between the examinees item performance and the set of traits underlying the performance can be defined by an item characteristic curve (Hambleton & Swaminathan, 1985). Regardless of group membership, as the level of ability increases, the probability of a correct response to an item increases (Hambleton & Cook, 1977).

There are three models in Item Response Theory. Figure 1 presents the three-parameter model, made up of the item discrimination "a" parameter, the item difficulty "b" parameter, and the guessing "c" parameter (Warm, 1978). The item discrimination parameter indicates the slope of the item characteristic curve. The item difficulty parameter indicates the location on the ability (d) axis where the probability for answering correctly is .50. The guessing parameter is the probability that a correct response occurs solely by chance.

INSERT FIGURE 1 ABOUT HERE

Figure 2 presents the two-parameter model. Notice that the item characteristic curves are asymptotic to zero, considering the guessing parameter negligible. The one-parameter or Rasch model is presented in Figure 3. In the Rasch model both the guessing and item discrimination parameters are considered negligible (Hambleton & Swamination, 1985). The present paper will focus on the calculations involved in the Rasch or one-parameter model.

INSERT FIGURES 2 AND 3 ABOUT HERE

The purpose of the Rasch model is to analyze differences in test scores that initially are not linear (Wright & Stone, 1979). To analyze these differences, data must be transformed into measures that are approximately linear. To achieve approximate linearity, probabilities are converted into logits, as presented in Table 1. Figure 4 presents a graph of the initial probabilities and a graph of the logit transformations of the probabilities. The transformation of probabilities to logits allows researchers to compare item difficulties and person abilities across tests (Warm, 1978).

INSERT TABLE 1 AND FIGURE 4 ABOUT HERE

The Rasch Model begins with a matrix of all items by persons, as presented in Table 2. Rows are persons, while columns are items. Within the matrix, a 1 denotes a correct response, while a 0 denotes an incorrect response. The final column presents the proportion of correct responses to the total number of responses for each person, while the final row presents the proportion of correct response to the total number of responses for each item. Next, as seen in Table 3, the people and items with all correct or incorrect responses are removed. No estimation can be obtained for these persons and items, because the data contain no information about these item difficulties or person abilities. Notice in Table 3 that items 18, 1, 2, and 3 and persons 35 and 36 are not included. Item 18 contains no information because no person answered it correctly; therefore, it would be impossible to estimate how difficult the item really is given only these data. Likewise, person 35 answered no items correctly, leaving no way to assess with the available data the ability for this person. When item 18 was omitted, person 36 was left with a perfect score. Therefore, person 36 had to be eliminated. Person 35 was omitted for missing all the items, leaving items 1, 2, and 3 with all correct responses. Therefore, these items had to be eliminated also. After eliminating these items and persons, new proportions are calculated using the remaining 34 people and 14 items displayed in Table 3.

INSERT TABLES 2 AND 3 ABOUT HERE

The next step in the Rasch model is to calibrate the initial item difficulties, as presented in Table 4. Item scores are listed in descending order by the number of correct responses and then by the frequency of their occurrence. Then the proportions are converted into logits. Logits are calculated by taking the natural log of the ratio of the proportion incorrect divided by the proportion correct. Once the proportions are transformed into logits the mean and variance for each distribution is computed. The mean (Avg) is then used to center the item logits at zero and the variance

(U) will be used in computing final calibrations. Notice that logits (d) are no longer bounded by zero and one, but have been transformed to a new scale that is infinite in both directions and is approximately linear to the underlying variable.

INSERT TABLE 4 ABOUT HERE

Once the initial item difficulties are calibrated, the initial person abilities are calibrated, as presented in Table 5. First, the possible correct answers for items are listed in ascending order, along with their associated frequencies. Then, the natural log of the proportion of successes is divided by the proportion of failures to convert the proportions into logits. The mean (ydot) and variance (V) are then calculated. The variance will be used in calculating the expansion factors for final calibrations.

INSERT TABLE 5 ABOUT HERE

To reach the final estimates for item difficulties and person abilities expansion factors are applied to the original estimates. The purpose of the expansion factor is to remove the effect of sample spread and test width to give final estimates that are neither person dependent or item dependent.

The formula for the expansion factor for person abilities due to test width is: SQRT((1+(U/2.89))/(1-((U*V)/8.35))), where U from Table 4 is the variance for item difficulties and V from Table 5 is the variance for person abilities. The formula for the expansion factor due to sample spread is: SQRT((1+(V/2.89))/(1-((V*U)/8.35))). In Table 6 the sample spread expansion factor is multiplied by the initial item calibration to yield the corrected item calibration. Likewise, in Table 7 the test width expansion factor is applied to the initial person measure to yield the corrected or final person ability measure.

INSERT TABLES 6 AND 7 ABOUT HERE

The Rasch model does not end with the final estimates of item difficulty and person ability. The fit of the model to the data must be evaluated (Hambleton & Cook, 1977), and not simply assumed. This is done by observing the differences between estimates of ability and difficulty for each person and item. Table 8 is a matrix of the responses of the 34 persons to 14 items. The last row presents the item difficulties, while the last column presents the person abilities. The double line in the table represents the point where person ability equals item difficulty. In theory, all the responses to the left or below the double line should be correct. Likewise, all the responses to the right or above the line should be incorrect. Answers not fitting with the theory are considered aberrations. In Table 8 these aberrations are underlined. For example, person four has two aberrant responses: item 4 and item 7. Item 14 has three aberrant responses: persons 23, 34, and 15.

INSERT TABLE 8 ABOUT HERE

Once the aberrations are identified, a fit analysis is computed for individual persons and items. Table 9 is an example of a fit analysis for person 19. The line between item 10 and item 11 represents the point where the person ability, 0.357, is equal to item difficulty, between .0375 and 1.174. According to the model, everything to the left of the line should be correct, denoted 1. Everything to the right of the line should be incorrect, denoted 0. There are four responses that do not fit the model: items 6, 9, 10, and 13. These aberrations are underlined. To compute the fit analysis the difference between the person ability and each item difficulty is first calculated. Next, a z^2 is calculated for each aberrant item using the formula: $z^2 = \exp|b-d|$. The variance (V) is then calculated by dividing the sum of the z^2 values by the number of items minus one (v-1). The variance is used to calculate a t-statistic using the formula: $t_{(df=v-1)} =$ ((ln(V))=(V-1)))*(((v-1)/8)**.5). For example, the calculated t-value for person 19 is 2.24, compared with the critical t-value at alpha=0.05 which is 2.160. Therefore, person 19 is not consistent with the model and should be removed from the data.

INSERT TABLE 9 ABOUT HERE

Not only is the fit analysis calculated for persons, but it is also calculated for items. Table 10 is an example of a fit analysis for item 8. The line between person 27 and person 11 represents the point where item difficulty, -1.836, is equal to person ability, between -1.973 and -1.266. Theoretically, everything above the line should be incorrect, while everything below should be correct. There are eight responses that are aberrant: persons 33, 27, 11, 12, 9, 29, 31, and 34. A z^2 is then calculated for each aberrant person using the same formula that was used for person fit analysis. The z^2 values are summed and divided by the number of people minus one (n-1) to calculate the variance. A t-statistic is then calculated using the formula: $t_{(df=n-1)}$= ((ln(V))=(V-1)))*(((n-1)/8)**.5). For example, the calculated t-value for item 8 is 3.65 compared with the critical t-value at alpha=0.05 which is 2.042. Therefore, item 8 is not consistent with the model and should be removed from the data. In fact, all items and persons found to be statistically significant are removed from the data and the entire analysis is repeated from the remaining score distributions until no items or persons are statistically significant.

INSERT TABLE 9 ABOUT HERE

To test whether the final calibrations are truly group independent, researchers may choose to do a cross validation. By tradition, this is typically done by dividing persons in a large sample with a large spread into six ability groupings. Item calibrations are then computed separately for each group. If the item calibrations for the total sample are similar to the six separate sets of item calibrations, then there is evidence that the final calibrations are sample independent.

The group-dependence and test-dependence of the classical measurement models have limited the appropriateness of comparing items and persons across tests. But, with IRT and the Rasch Model, item difficulties and person abilities can now be compared linearly, free of group and test dependence, if the IRT model fits the data.

However, Lawson (1991) has raised concerns about how often this occurs. Lawson (1991) analyzed the differences between the classical measurement model and the Rasch model to evaluate the benefits of using the Rasch model. The analysis revealed that both procedures, classical and Rasch yielded almost perfectly correlated results as regards to both person abilities and item difficulties. These similarities are obscured only because IRT models express both person abilities and item difficulties in logits, which are units unfamiliar to some people.

In the present paper, to analyze the differences between the Rasch calibrations and the classical measurement calibrations, a regression analysis was performed to find both a correlation between the two measures and to plot a scatterplot of the two measures with their regression line. Table 10 presents the item probabilities from Table 3 and the item difficulties from Table 6. The probabilities and difficulties were correlated using a regression analysis which revealed an r = -.985. This supports Lawson's analysis that the two sets of statistics are <u>very</u> highly correlated.

INSERT TABLE 11 ABOUT HERE

Table 11 presents the number of items correct from Table 3 and the person abilities from Table 7. The number correct and the person abilities were correlated using a regression analysis revealing an r = .997. Again, the two sets of statistics are <u>very</u> highly correlated. Figures 5 and 6 present scatterplots of the item probabilities and item difficulties and the number correct and the person abilities and their associated regression lines. Again, this confirms Lawson's claim that the two sets of statistics are almost identical.

INSERT TABLE 12 AND FIGURES 5 AND 6 ABOUT HERE

The results from Lawson's chapter and the present paper challenge the idea that Rasch latent trait measurement is superior to classical measurement because its estimates are item-free and sample-free. The high correlations between the two measures can be explained by only one of two possibilities: (1) the calibrations in the Rasch model are not truly item-free and sample-free, or (2) the calibrations in the classical measurement model are also item-free and sample-free. Though Rasch model procedures may superior in other ways (Lawson, 1991), the superiority does not arise from unique item-free and sample-free calibrations.

References

Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 14, 75-96.

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

Lawson, S. (1991). One parameter latent trait measurement: Do the results justify the effort? in B. Thompson (Ed.), Advances in education research: Substantive findings, methodological developments (Vol. 1, pp. 159-168). Greenwich, CT: JAI Press. (International Book Sellers Number: 1-55938-316-X).

Warm, T. A. (1978). A primer of item response theory (Technical Rep. No. 940279). U. S. Coast Guard Institute, Oklahoma City, OK: U. S. Government Printing Office.

Wright, B. D., & Stone, M. H. (1979). Best test design: Rasch measurement. Chicago: Mesa.

Table 1

Transformations of Probability Proportions to Logits

p	q	(1-p)/p	ln(1-p)/p
0.01	0.99	99.00	4.60
0.05	0.95	19.00	2.94
0.10	0.90	9.00	2.20
0.15	0.85	5.67	1.73
0.20	0.80	4.00	1.39
0.25	0.75	3.00	1.10
0.30	0.70	2.33	0.85
0.35	0.65	1.86	0.62
0.40	0.60	1.50	0.41
0.45	0.55	1.22	0.20
0.50	0.50	1.00	0.00
0.55	0.45	0.82	-0.20
0.60	0.40	0.67	-0.41
0.65	0.35	0.54	-0.62
0.70	0.30	0.43	-0.85
0.75	0.25	0.33	-1.10
0.80	0.20	0.25	-1.39
0.85	0.15	0.18	-1.73
0.90	0.10	0.11	-2.20
0.95	0.05	0.05	-2.94
0.99	0.01	0.01	-4.60

Table 2. Ordered Responses of 36 Persons to 18 Items

Person Name	Item Name: 1	2	3	4	5	7	6	9	8	10	11	13	12	14	15	16	17	18	Person Score	P of 18
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00
25	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	5	0.28
4	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	5	0.28
33	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	6	0.33
1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	6	0.33
27	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	7	0.39
11	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	8	0.44
12	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	8	0.44
17	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	8	0.44
19	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	8	0.44
30	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	9	0.50
2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	9	0.50
3	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	9	0.50
5	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	9	0.50
6	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	9	0.50
8	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	9	0.50
9	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	9	0.50
13	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10	0.56
16	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10	0.56
26	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10	0.56
28	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	10	0.56
29	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	10	0.56
31	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	10	0.56
10	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10	0.56
18	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	10	0.56
14	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	10	0.56
32	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	11	0.61
20	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	12	0.67
21	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	12	0.67
22	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	13	0.72
23	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	13	0.72
34	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	13	0.72
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	14	0.78
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	15	0.83
24	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	15	0.83
36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	17	0.94
Item Score	35	35	35	33	33	31	27	27	26	24	14	10	7	5	3	2	2	0		
P of 36	0.97	0.97	0.97	0.92	0.92	0.86	0.75	0.75	0.72	0.67	0.39	0.28	0.19	0.14	0.08	0.06	0.06	0		

17

BEST COPY AVAILABLE

Table 3. Edited and Ordered Responses of 34 Persons to 14 Items

Person Name	4	5	7	6	9	8	10	11	13	12	14	15	16	17	Person Score	P of 14
25	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2	0.143
4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2	0.143
33	1	0	0	0	1	1	0	0	0	0	0	0	0	0	3	0.214
1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	3	0.214
27	1	0	0	1	1	1	0	0	0	0	0	0	0	0	4	0.286
11	1	1	1	1	1	0	0	0	0	0	0	0	0	0	5	0.357
12	1	1	1	0	1	0	1	0	0	0	0	0	0	0	5	0.357
17	1	1	1	0	0	1	1	0	0	0	0	0	0	0	5	0.357
19	1	1	1	0	0	1	0	0	1	0	0	0	0	0	5	0.357
30	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0.429
2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0.429
3	1	1	1	1	0	1	1	0	0	0	0	0	0	0	6	0.429
5	1	1	1	0	1	1	1	0	0	0	0	0	0	0	6	0.429
6	1	1	1	0	1	1	1	0	0	0	0	0	0	0	6	0.429
8	1	1	1	1	0	1	1	0	0	0	0	0	0	0	6	0.429
9	1	1	1	1	0	0	1	1	0	0	0	0	0	0	6	0.429
13	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.5
16	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.5
26	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.5
28	1	1	1	0	1	1	1	1	0	0	0	0	0	0	7	0.5
29	1	1	1	1	1	0	1	1	0	0	0	0	0	0	7	0.5
31	1	1	1	1	1	0	1	1	0	0	0	0	0	0	7	0.5
10	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.5
18	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0.5
14	1	1	1	1	1	1	0	0	1	0	0	0	0	0	7	0.5
32	1	1	1	1	1	1	1	0	0	1	0	0	0	0	8	0.571
20	1	1	1	1	1	1	1	1	1	0	0	0	0	0	9	0.643
21	1	1	1	1	1	1	1	1	0	1	0	0	0	0	9	0.643
22	1	1	1	1	1	1	1	1	1	1	0	0	0	0	10	0.714
23	1	1	1	1	1	1	1	1	1	0	1	0	0	0	10	0.714
34	1	1	1	1	1	0	1	1	1	1	1	0	0	0	10	0.714
15	1	1	1	1	1	1	1	1	1	1	0	1	0	0	11	0.786
7	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12	0.857
24	1	1	1	1	1	1	1	1	1	0	1	0	1	1	12	0.857
Item Score	32	32	30	26	26	25	23	13	9	6	4	2	1	1		
P of 34	0.941	0.941	0.882	0.765	0.765	0.735	0.676	0.382	0.265	0.176	0.118	0.059	0.029	0.029		

Table 4. Initial Calibration of Item Difficulties

Item #	Item Score	freqi	Pi	1-Pi	1-Pi/ Pi	logit ratio = xi	freqi x logit	xi square	freqi xi sq	d=logit -mean	fi x sq mean
4,5	32	2	0.941	0.059	0.063	-2.773	-5.545	7.687	15.374	-2.997	0.101
7	30	1	0.882	0.118	0.133	-2.015	-2.015	4.060	4.060	-2.240	0.050
6,9	26	2	0.765	0.235	0.308	-1.179	-2.357	1.389	2.778	-1.403	0.101
8	25	1	0.735	0.265	0.360	-1.022	-1.022	1.044	1.044	-1.246	0.050
10	23	1	0.676	0.324	0.478	-0.738	-0.738	0.544	0.544	-0.962	0.050
11	13	1	0.382	0.618	1.615	0.480	0.480	0.230	0.230	0.255	0.050
13	9	1	0.265	0.735	2.778	1.022	1.022	1.044	1.044	0.797	0.050
12	6	1	0.176	0.824	4.667	1.540	1.540	2.373	2.373	1.316	0.050
14	4	1	0.118	0.882	7.500	2.015	2.015	4.060	4.060	1.790	0.050
15	2	1	0.059	0.941	16.000	2.773	2.773	7.687	7.687	2.548	0.050
16,17	1	2	0.029	0.971	33.000	3.497	6.993	12.226	24.451	3.272	0.101
	v=	14				Sums	3.146		63.646		0.707
						Avg	0.225				
						Avg2	0.050				

U = Sumfixi2 - Sumfiavg2 · / v-1

U = 63.646 0.707 13

U = 62.939 13

U = 4.841

BEST COPY AVAILABLE

Table 5. Initial Calibration of Person Abilities

Possible right	fr	Score prop	1 - Score prop	ScrPr/ 1 - ScrPr	logit ratio = yr	freqr x logit	yr square	freqr yr sq	initial person measure
1	0	0.071	0.929	0.077	-2.565	0.000	6.579	0.000	-2.565
2	2	0.143	0.857	0.167	-1.792	-3.584	3.210	6.421	-1.792
3	2	0.214	0.786	0.273	-1.299	-2.599	1.688	3.376	-1.299
4	1	0.286	0.714	0.400	-0.916	-0.916	0.840	0.840	-0.916
5	4	0.357	0.643	0.556	-0.588	-2.351	0.345	1.382	-0.588
6	7	0.429	0.571	0.750	- .288	-2.014	0.083	0.579	-0.288
7	9	0.500	0.500	1.000	0.000	0.000	0.000	0.000	0.000
8	1	0.571	0.429	1.333	0.288	0.288	0.083	0.083	0.288
9	2	0.643	0.357	1.800	0.588	1.176	0.345	0.691	0.588
10	3	0.714	0.286	2.500	0.916	2.749	0.840	2.519	0.916
11	1	0.786	0.214	3.667	1.299	1.299	1.688	1.688	1.299
12	2	0.857	0.143	6.000	1.792	3.584	3.210	6.421	1.792
13	0	0.929	0.071	13.000	2.565	0.000	6.579	0.000	2.565
n =	34				Sums	-2.368		23.999	
		ydot=	-0.0697						
		V=	0.7294						

Table 6. Final Estimates of Item Difficulties

Item #	Initial Item Calibration	Sample Spread Expansion Factor	Corrected Item Calibration	Item score	Calibrated Standard Error
4,5	-2.997	1.473	-4.415	32	1.074
7	-2.240	1.473	-3.299	30	0.784
6,9	-1.403	1.473	-2.067	26	0.596
8	-1.246	1.473	-1.836	25	0.573
10	-0.962	1.473	-1.418	23	0.540
11	0.255	1.473	0.375	13	0.520
13	0.797	1.473	1.174	9	0.573
12	1.316	1.473	1.938	6	0.663
14	1.790	1.473	2.637	4	0.784
15	2.548	1.473	3.753	2	1.074
16,17	3.272	1.473	4.820	1	1.495

Table 7. Final Estimates of Person Measures

Possible Test Score	Initial Measure	Test Width Expansion Factor	Corrected Measure	Measure Standard Error
1	-2.565	2.153	-5.522	2.234
2	-1.792	2.153	-3.858	1.644
3	-1.299	2.153	-2.797	1.402
4	-0.916	2.153	-1.973	1.274
5	-0.588	2.153	-1.266	1.201
6	-0.288	2.153	-0.619	1.163
7	0.000	2.153	0.000	1.151
8	0.288	2.153	0.619	1.163
9	0.588	2.153	1.266	1.201
10	0.916	2.153	1.973	1.274
11	1.299	2.153	2.797	1.402
12	1.792	2.153	3.858	1.6[illegible]4
13	2.565	2.153	5.522	2.234

Table 8. Aberrant Responses of 34 people to 14 items

Person Name	Easy Items 4	5	7	6	9	8	10	Hard Items 11	13	12	14	15	16	17	Person Score	P of 14	Score Ability
25	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2	0.1429	-3.858
4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2	0.1429	-3.858
33	1	0	0	0	1	1	0	0	0	0	0	0	0	0	3	0.2143	-2.797
1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	3	0.2143	-2.797
27	1	0	0	1	1	1	0	0	0	0	0	0	0	0	4	0.2857	-1.973
11	1	1	1	1	1	0	0	0	0	0	0	0	0	0	5	0.3571	-1.266
12	1	1	1	0	1	0	1	0	0	0	0	0	0	0	5	0.3571	-1.266
17	1	1	1	0	0	1	1	0	0	0	0	0	0	0	5	0.3571	-1.266
19	1	1	1	0	0	1	0	0	1	0	0	0	0	0	5	0.3571	-1.266
30	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0.4286	-0.619
2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0.4286	-0.619
3	1	1	1	1	0	1	1	0	0	0	0	0	0	0	6	0.4286	-0.619
5	1	1	1	0	1	1	1	0	0	0	0	0	0	0	6	0.4286	-0.619
6	1	1	1	0	1	1	1	0	0	0	0	0	0	0	6	0.4286	-0.619
8	1	1	1	1	0	1	1	0	0	0	0	0	0	0	6	0.4286	-0.619
9	1	1	1	1	0	0	1	1	0	0	0	0	0	0	6	0.4286	-0.619
13	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.5000	0
16	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.5000	0
26	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.5000	0
28	1	1	1	0	1	1	1	1	0	0	0	0	0	0	7	0.5000	0
29	1	1	1	1	1	0	1	1	0	0	0	0	0	0	7	0.5000	0
31	1	1	1	1	1	0	1	1	0	0	0	0	0	0	7	0.5000	0
10	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.5000	0
18	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0.5000	0
14	1	1	1	1	1	1	0	0	1	0	0	0	0	0	7	0.5000	0
32	1	1	1	1	1	1	1	0	0	1	0	0	0	0	8	0.5714	0.619
20	1	1	1	1	1	1	1	1	1	0	0	0	0	0	9	0.6429	1.266
21	1	1	1	1	1	1	1	1	0	1	0	0	0	0	9	0.6429	1.266
22	1	1	1	1	1	1	1	1	1	1	0	0	0	0	10	0.7143	1.973
23	1	1	1	1	1	1	1	1	1	0	1	0	0	0	10	0.7143	1.973
34	1	1	1	1	1	0	1	1	1	1	1	0	0	0	10	0.7143	1.973
15	1	1	1	1	1	1	1	1	1	1	0	1	0	0	11	0.7857	2.797
7	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12	0.8571	3.858
24	1	1	1	1	1	1	1	1	1	0	1	0	1	1	12	0.8571	3.858
Item Score	32	32	30	26	26	25	23	13	9	6	4	2	1	1			
P of 34	0.9412	0.9412	0.8824	0.7647	0.7647	0.7353	0.6765	0.3824	0.2647	0.1765	0.1176	0.0588	0.0294	0.0294			
Diff	-4.415	-4.415	-3.299	-2.067	-2.067	-1.836	-1.418	0.375	1.174	1.938	2.637	3.753	4.82	4.82			

29

BEST COPY AVAILABLE

Table 9. Fit analysis for Illustrative Person

Item													
Person	4	5	7	6	9	8	10	11	13	12	14	16	17
19	1	1	1	0	0	1	0	0	1	0	0	0	0
Person Calibration 0.357													
Pi	0.941	0.941	0.882	0.765	0.765	0.735	0.677	0.384	0.265	0.177	0.118	0.029	0.029
Item Calib	-4.42	-4.42	-3.3	-3.3	-2.07	-1.84	-1.42	0.0375	1.174	1.938	2.637	3.753	4.82
x=0 (b-d)				2.427	2.197		0.3195	-0.817		-2.28	-3.396	-4.463	-4.463
x=1 (d-b)	4.777	4.777	3.657			1.777			-1.581				
Abber				2.427	2.197		0.3195		-1.581				
z2				11.3249	8.99798		1.37644		4.85981				
z2=exp|abberance|													
V = SOS / v-1													
	26.5591	13											
	2.04301												
t(df=v-1) = ((ln(V)) + (V-1)) * (((v-1) / 8)**.5)													
		ln	2.04301	2.04301	1	14	1	8 **.5					
			0.71442	2.04301	1	14	1	8 **.5					
			0.71442	1.04301		14	1	8 **.5					
				1.75743		14	1	8 **.5					
				1.75743		13		8 **.5					
				1.75743			1.625	**.5					
				1.75743			1.27475						
			t=		2.24029								

BEST COPY AVAILABLE

Table 10. Fit Analysis for Illustrative Item

Item 8
Item
Calibration
-1.836

Person	Person Score	br	Item Score	x=0 b-d	x=1 d-b	Aberrant	z2
25	2	-3.86	0	-2.02			
4	2	-3.86	0	-2.02			
33	3	-2.8	1		-0.96	-0.961	2.614
1	3	-2.8	0	-0.96			
27	4	-1.97	1		-0.14	-0.137	1.147
11	5	-1.27	0	0.57		0.57	1.768
12	5	-1.27	0	0.57		0.57	1.768
17	5	-1.27	1		0.57		
19	5	-1.27	1		0.57		
30	6	-0.62	1		1.217		
2	6	-0.62	1		1.217		
3	6	-0.62	1		1.217		
5	6	-0.62	1		1.217		
6	6	-0.62	1		1.217		
8	6	-0.62	1		1.217		
9	6	-0.62	0	1.217		1.217	3.377
13	7	0	1		1.836		
16	7	0	1		1.836		
26	7	0	1		1.836		
28	7	0	1		1.836		
29	7	0	1		1.836		
31	7	0	0	1.836		1.836	6.271
10	7	0	0	1.836		1.836	6.271
18	7	0	1		1.836		
14	7	0	1		1.836		
32	8	0.619	1		2.455		
20	9	1.266	1		3.102		
21	9	1.266	1		3.102		
22	10	1.973	1		3.809		
23	10	1.973	1		3.809		
34	10	1.973	0	3.809		3.809	45.11
15	11	2.797	1		4.633		
7	12	3.858	1		5.694		
24	12	3.858	1		5.694		
						SOS	68.32

V = SOS / n-1
68.32 33

2.07

t(df=n-1) = ((ln(V)) + (V-1)) * (((n-1) / 8)**.5)

ln	2.07	2.07	1	34	1	8	**.5
	0.728	2.07	1	34	1	8	**.5
	0.728	1.07		34	1	8	**.5
		1.798		34	1	8	**.5
		1.798		33		8	**.5
		1.798			4.125		**.5
		1.798			2.031		
		t=	3.652				

Table 11. Item probabilities and difficulties

Item	Probability	Difficulty
4	0.941176471	-4.415
5	0.941176471	-4.415
7	0.882352941	-3.299
6	0.764705882	-2.067
9	0.764705882	-2.067
8	0.735294118	-1.836
10	0.676470588	-1.418
11	0.382352941	0.375
13	0.264705882	1.174
12	0.176470588	1.938
14	0.117647059	2.637
15	0.058823529	3.753
16	0.029411765	4.820
17	0.029411765	4.820

Table 12. Number of items correct and person abilities.

Person	#correct	Ability
25	2	-3.858
4	2	-3.858
33	3	-2.797
1	3	-2.797
27	4	-1.973
11	5	-1.266
12	5	-1.266
17	5	-1.266
19	5	-1.266
30	6	-0.619
2	6	-0.619
3	6	-0.619
5	6	-0.619
6	6	-0.619
8	6	-0.619

Figure Captions

Figure 1. Three-parameter item characteristic curves.

Figure 2. Two-parameter item characteristic curves.

Figure 3. One-parameter item characteristic curves.

Figure 4. Graph of probability proportions and probability proportions transformed into logits.

Figure 5. Scatterplot of item probabilities with item difficulties.

Figure 6. Scatterplot of number correct with person abilities.

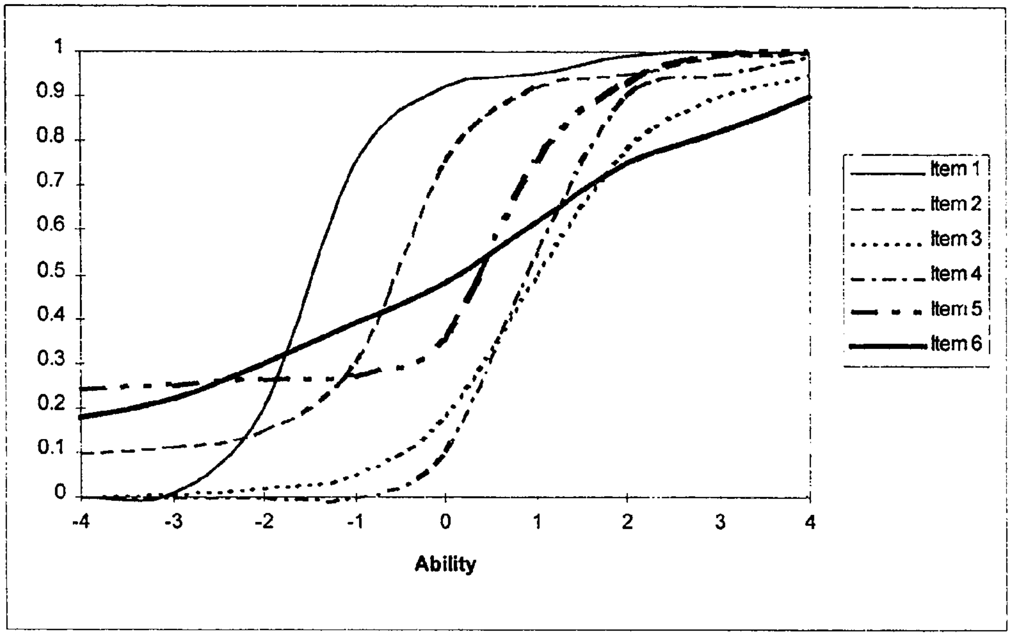

Figure 1. Three-parameter item characteristic curves.

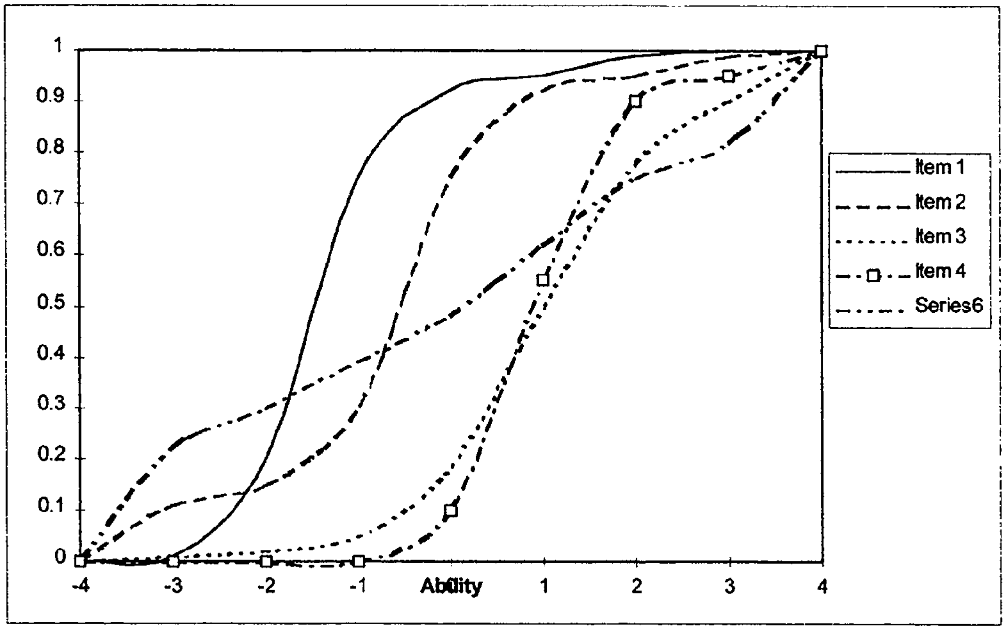

Figure 2. Two-parameter item characteristic curves.

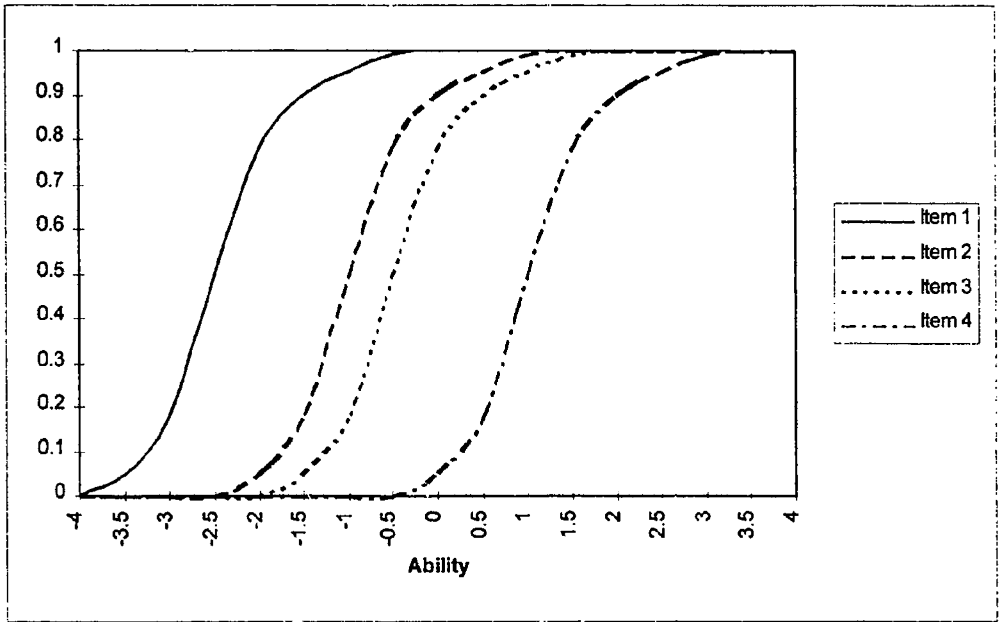

Figure 3. One-parameter item characteristic curves.

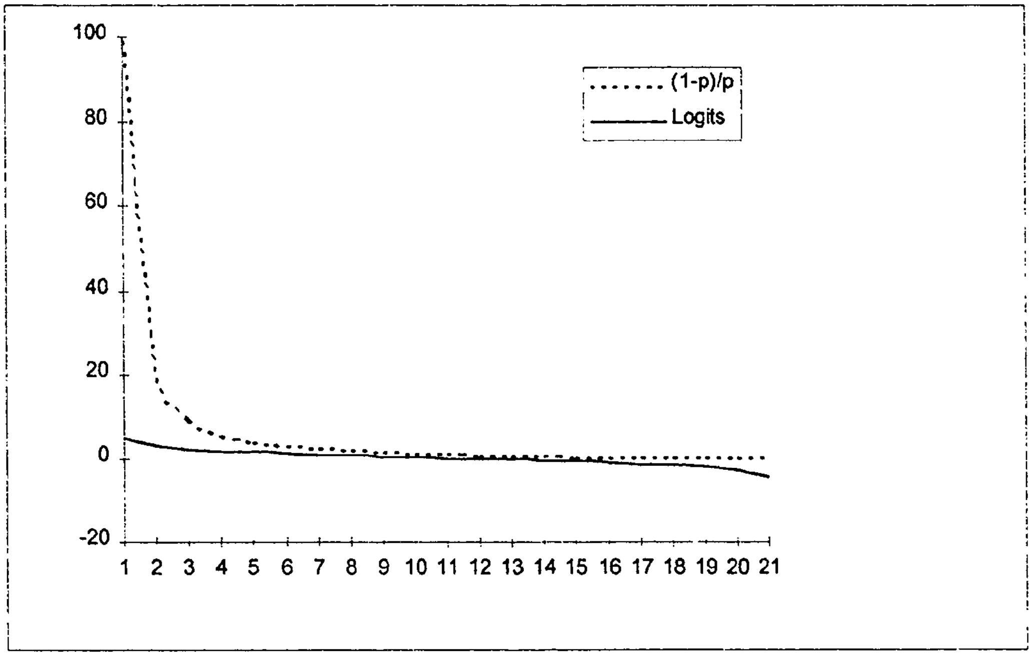

Figure 4. Graph of probability proportions and probability proportions transformed into logits.

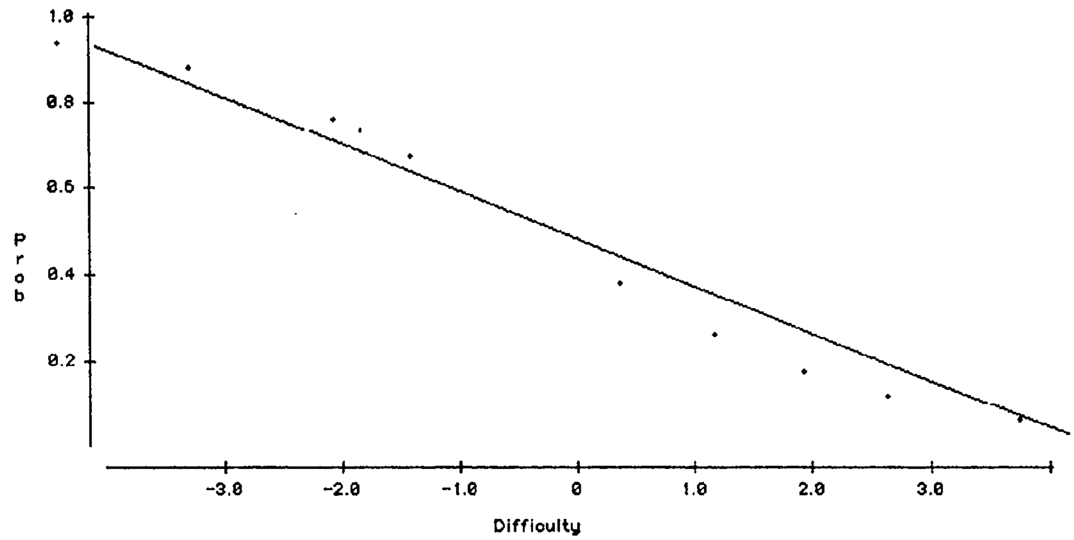

Figure 5. Scatterplot of item probabilities with item difficulties.

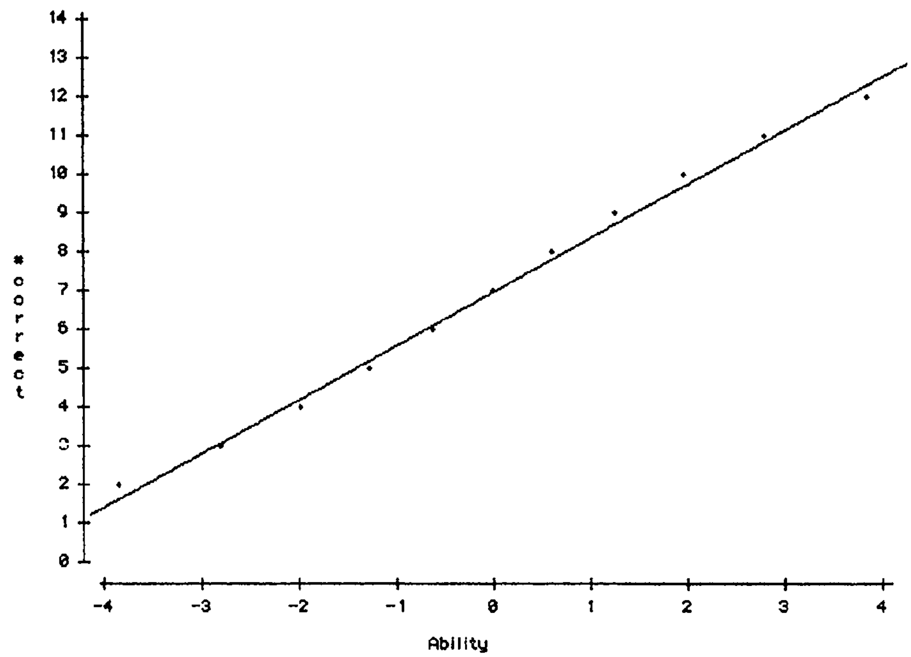

Figure 6. Scatterplot of number correct with person abilities.