An NCME Instructional Module on

Comparison of 1-, 2-, and 3-Parameter IRT Models

Deborah Harris
American College Testing Program

This module discusses the 1-, 2-, and 3-parameter logistic item response theory models. Mathematical formulas are given for each model, and comparisons among the three models are made. Figures are included to illustrate the effects of changing the a, b, or c parameter, and a single data set is used to illustrate the effects of estimating parameter values (as opposed to the true parameter values) and to compare parameter estimates achieved though applying the different models. The estimation procedure itself is discussed briefly. Discussions of model assumptions, such as dimensionality and local independence, can be found in many of the annotated references (e.g., Hambleton, 1988).

Item response theory (IRT) attempts to model the relationship between an unobserved variable, usually conceptualized as an examinee's ability, and the probability of the examinee correctly responding to any particular test item. Three currently used models are the 3-parameter logistic, the 2-parameter logistic, and the 1-parameter logistic models. (The 1-parameter model is referred to as the Rasch model.) These models all assume a single underlying ability—usually a continuous, unbounded variable designated as θ—for examinees, but vary in the characteristics they ascribe to items. All three models have an item difficulty parameter *(b)*, which is the point of inflection on the ability (θ) scale. For the 1-parameter and 2-parameter models, this is the point on the ability (θ) scale at which an examinee has a 50% probability of correctly answering an item. For the 3-parameter item, this is the point at which the probability of correctly answering an item is $(1 + c)/2$, where c is a lower asymptote parameter (see the discussion that follows). Theoretically, difficulty values can range from $-\infty$ to $+\infty$; in practice, values usually are in the range of -3 to $+3$, when θ is scaled to have a mean of 0 and standard deviation of 1.0. Similarly, examinee ability values can range from $-\infty$ to $+\infty$, but values in excess of ± 3 seldom are seen. Items with high values of b are *hard* items, with low-ability examinees having low probabilities of correctly responding to the item. Items with low values of b are *easy* items, with most examinees, even those with low-ability values, having at least a moderate probability of answering the item correctly.

The 3- and 2-parameter models also have a discrimination parameter *(a)* that allows items to differentially discriminate among examinees. Technically, a is defined as the slope of the item characteristic curve (ICC) at the point of inflection (see Baker, 1985, p. 21). The a parameter can range in value from $-\infty$ to $+\infty$, with typical values being ≤ 2.0 for multiple-choice items. The higher the a value, the more sharply the item discriminates between examinees at the point of inflection.

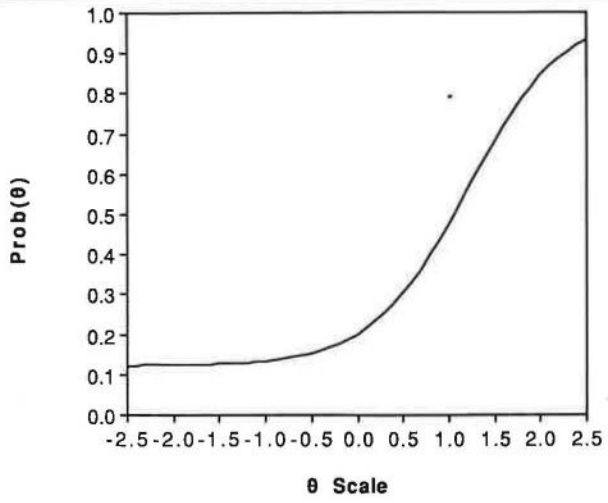
The 3-parameter model also has a lower asymptote parameter *(c)*, which is sometimes referred to as the pseudochance parameter. This parameter allows for examinees, even ones with low ability, to have perhaps substantial probability of correctly answering even moderate or hard items. Theoretically, c ranges from 0.0 to 1.0, but is typically < 0.3.

Each of the models uses a logistic function to relate examinee ability and the item parameter(s) to the probability of correctly responding to an item. For each item, an item characteristic curve can be constructed that plots the probability of correctly responding to an item, given the θ level. This relation generally will be nonlinear, as the θ variable is unbounded and probability is bounded. Examples of item characteristic curves for each of the three logistic models are shown in Figure 1.

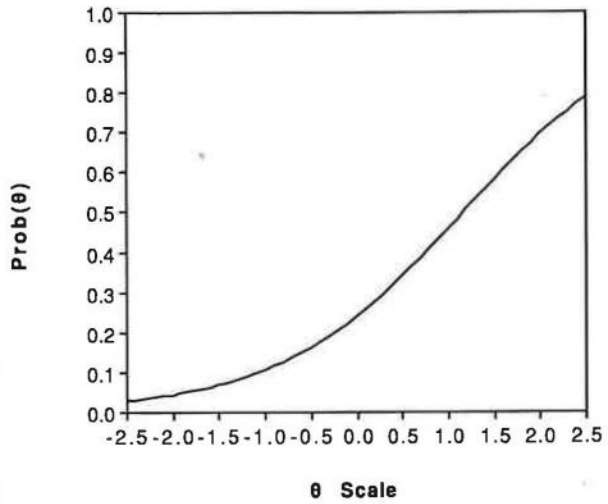
Deborah Harris is a Psychometrician, American College Testing Program, P.O. Box 168, Iowa City, IA 52243. She specializes in psychometrics.

Series Information

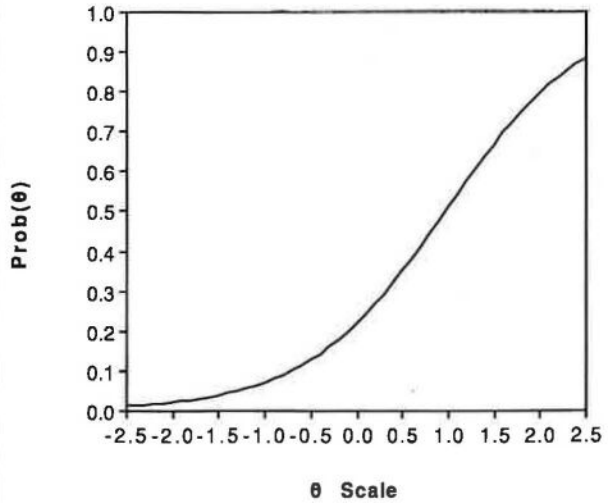
ITEMS is a series of units designed to facilitate instruction in educational measurement. These units are published by the National Council on Measurement in Education. Fred Brown, Iowa State University, has served as the editor for this module.



(a) *3-parameter logistic model*



(b) *2-parameter logistic model*



(c) *1-parameter logistic model*

FIGURE 1. *Examples of item characteristic curves derived from the 3-, 2- and 1-parameter logistic models*

Define the three parameters used to describe items in the 1-, 2-, and 3-parameter models.

The 3-Parameter Model

The 3-parameter model is defined mathematically as:

$$P(\theta) = c + \frac{(1 - c)}{1 + \exp[-1.7a(\theta - b)]}, \qquad (1)$$

where a, b, and c are as defined above.

The effects of varying different parameters in a 3-parameter logistic model are illustrated in Figures 2 through 4. In Figure 2, the three item characteristic curves plotted vary only in their a (discrimination) parameter. Notice that the higher the value of a, the steeper the curve, thus the more sharply the item discriminates between examinees. When a is higher, however, the item discriminates over a smaller range of θs. A test constructer would need to decide whether items that sharply discriminate over a narrow range or items that are less discriminating but operate over a broader range are more desirable for a particular purpose.

Figure 3 shows three item characteristic curves that vary only in the b parameter. Notice that because the a parameter is constant, the discrimination power is constant. However, the point on the ability scale where the item discriminates best varies.

In Figure 4, the c parameter varies from 0.0 to .25 with a and b held constant. Notice that the lower tails of the three item characteristic curves differ.

Table 1 shows item response data (coded 0 = incorrect; 1 = correct) for 10 examinees and 14 items. Using the full data set of 3000 examinees, from which the 10 examinees in Table 1 were extracted, the item characteristic curve in Figure 5 was constructed for one of the items, using a 3-parameter logistic model. Based on the item response data alone, would it be possible to identify which item was used in Figure 5? Probably not; however, knowing the a, b, and c parameter values for each of the items would allow one to identify the item used.

The 14 items in Table 1 have the following parameter values:

Item	a	b	c
1	1.36	−1.34	.20
2	.88	.05	.20
3	.64	.28	.12
4	1.60	.98	.12
5	1.36	.28	.05
6	.64	−.88	.20
7	.88	.28	.05
8	1.12	.75	.12
9	1.60	.52	.27
10	1.36	1.21	.05
11	1.12	1.21	.12
12	1.12	.05	.27
13	.64	.75	.05
14	.64	.05	.20

Using this information, try to determine what item's characteristic curve is plotted in Figure 5.

Noting that the point of inflection *(b)* is slightly greater than +1.00, that the lower asymptote *(c)* is slightly above .10, and that the slope at the point of inflection is rather steep, one could deduce that item 11 is the one plotted. The item characteristic curve was constructed by inserting the given parameter values into Equation 1 for various θ values. To illustrate, for $\theta = 1.0$:

$$P(1.0) = .12 + \frac{(1.00 - .12)}{1.00 + \exp[-1.7(1.12)(1.00 - 1.21)]} = .47319,$$

which is the value plotted against $\theta = 1.0$ in Figure 5. Similar calculations were done for other values of θ.

The item characteristic curve shows the probability of responding correctly to an item for any given θ level. For item 11, the probability of an examinee with $\theta = 1.0$ correctly responding is .47. This means that (a) on average, in a group of 100 examinees with θs equal to 1.0, 47 will correctly respond to item 11, or (b) a given examinee with $\theta = 1.0$, when presented with 100 items similar to item 11, will answer correctly, on average, 47% of the items.

The 10 examinees used in Table 1 have θ values of −.05, .85, −.45, −1.50, 1.95, .51, .64, −.57, −.21, and −.47, respectively. Note that although examinee CC has a higher θ than examinee JJ, examinee JJ had a higher raw score. Although the θ level indicates the "amount of ability" an examinee has, it does not "predict" the examinee's test score without error. To illustrate how this can be, consider the probabilities of correctly answering a given item. The probability of an examinee with a θ of −.47 answering item 3 correctly is .38779; the probability of an examinee with a θ of −.45 correctly answering item 3 is .39186, and the probability of an examinee with a θ of .64 correctly answering item 3 is .64278. In this data set, however, examinee JJ (with a θ of −.47) responded correctly to item 3 whereas examinee CC (θ = −.45) and examinee GG (θ = .64) responded incorrectly. The data set in Table 1 was generated using known θ values. Results can be more anomalous when estimated ability values ($\hat{\theta}$s) are used instead of $\hat{\theta}$s.

The 2-Parameter Model

The 2-parameter logistic model is mathematically specified as:

$$P(\theta) = \frac{1}{1 + \exp[-1.7a(\theta - b)]}. \qquad (2)$$

Note that if c was set equal to zero in Equation 1, Equation 2 would result. Setting the c parameter to zero implies that an examinee of sufficiently low ability could have near zero probability of correctly responding to an item, which might be the case with supply rather than selection items.

Although the 2-parameter model is less general than the 3-parameter, in that it does not allow the lower asymptote to vary, the 2-parameter model sometimes is preferred because of the difficulties in estimating the c parameter in real data sets. Unlike the 3-parameter model, sufficient statistics

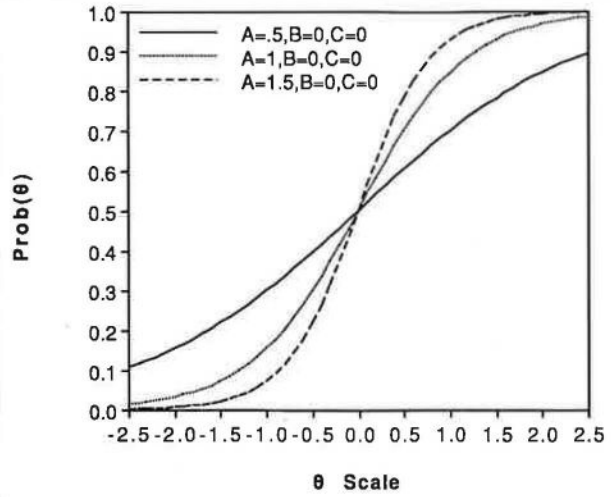


FIGURE 2. *ICCs derived by varying the a parameter in a 3-parameter logistic model*

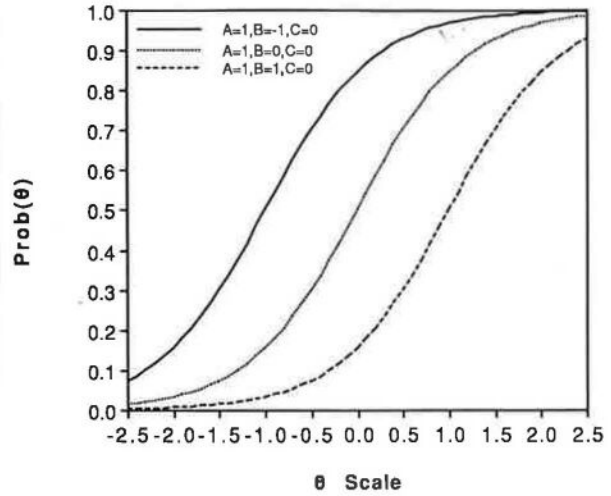


FIGURE 3. *ICCs derived by varying the b parameter in a 3-parameter logistic model*

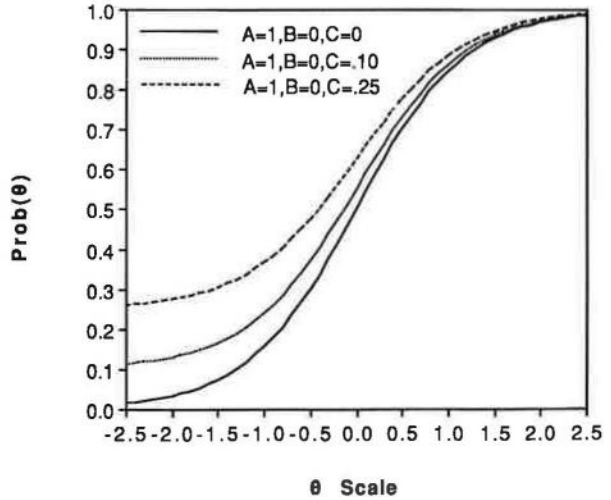


FIGURE 4. *ICCs derived by varying the c parameter in a 3-parameter logistic model*

TABLE 1

Item Responses for 10 Examinees to 14 Items

Examinee	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total raw score
AA	1	0	1	0	0	1	1	0	1	0	0	1	1	0	7
BB	1	1	1	1	1	1	1	0	1	1	0	0	1	0	10
CC	1	0	0	0	0	0	1	0	1	0	1	1	0	0	5
DD	1	0	1	0	0	0	0	0	0	0	0	1	0	0	3
EE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14
FF	1	0	1	0	1	1	1	1	1	0	0	1	0	1	9
GG	1	1	0	0	1	1	1	0	1	0	0	0	0	0	6
HH	1	0	0	0	0	1	1	0	1	0	0	0	0	0	4
II	1	0	0	0	0	1	0	0	1	0	0	0	0	1	4
JJ	1	1	1	0	1	0	0	1	1	0	0	1	0	0	7

do exist for calculating the parameters with the two parameter model.

The 1-Parameter Model

The 1-parameter model is mathematically specified as:

$$P(\theta) = \frac{1}{1 + \exp[-1.7a(\theta - b)]}, \quad (3)$$

where a is constant for all items (and is often scaled to equal one). Having a constant implies that all items on a test are equally discriminating. The items, however, may discriminate at different places on the ability scale; easy items discriminate among low-ability students as well as hard items discriminate among high-ability students.

The 1-parameter model has some advantages over the 2- and 3-parameter models: the total test score (with number right scoring) is a sufficient statistic for estimating θ, and the number of examinees correctly responding to an item is a sufficient statistic for estimating b. Thus, the model fits nicely with number-right scoring. Also, examinees with the same raw score will have the same $\hat{\theta}$, which is not the case with the 2- and 3-parameter models.

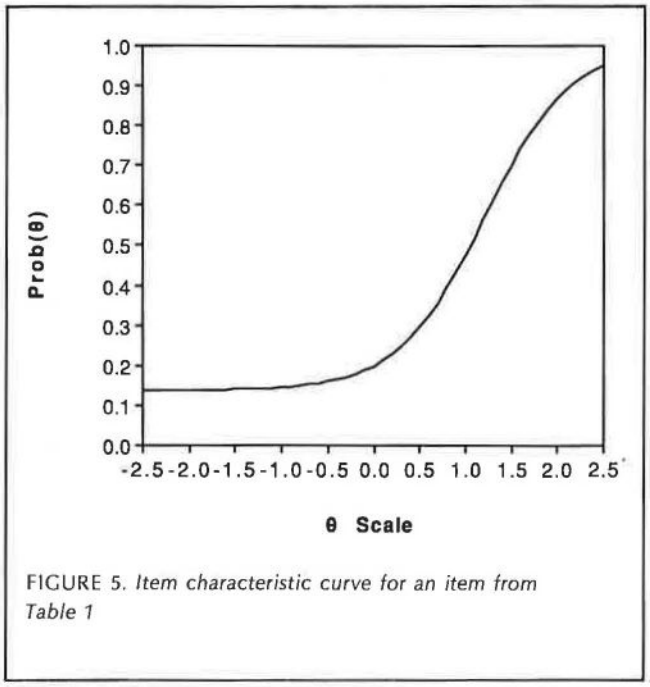


FIGURE 5. *Item characteristic curve for an item from Table 1*

Here, again, a choice must be made between generality (which parameters are allowed to vary) and simplicity. Although few, if any, tests meet the strict assumption of equal discrimination across items, the 1-parameter logistic model has found utility in a number of settings.

> Describe the major differences between the 1-, 2-, and 3-parameter models.

Estimation Procedures

In the discussion thus far, item parameters have been presented as *known;* in actual applications, both the item parameter(s) and the examinee parameters (θs) must be estimated using the examinees' responses to the items. In practical applications, these parameters are estimated by computer programs because of the vast number of parameters that must be estimated (1, 2, or 3 parameters for each item and 1 parameter for each examinee).

One popular estimation program is LOGIST (Wingersky, Barton, & Lord, 1982). LOGIST uses joint maximum likelihood procedures to estimate item and examinee parameters. It can be adapted to provide estimates for the 1-, 2-, and 3-parameter models. The item data are coded correct, incorrect, omitted, or not reached, and a Newton-Raphson iterative procedure is used. The particular values of a, b, and θ parameters will depend on the specific scale unit and origin chosen. LOGIST sets some restrictions on the ranges of a, c, and θ parameters and scales θ so that the mean and standard deviation (over a truncated range) will equal zero and one, respectively. The scaling of θ also sets the scale for the as and bs; cs are "scale free." This scaling is for each particular examinee group; parameters must be put on the same scale before comparing θs or item parameters from different computer runs (see Hambleton & Swaminathan, 1985, Chapter 10).

Estimation is a four stage procedure: (a) the θs and bs are estimated, then (b) as, bs, and cs, then (c) θs and bs, and (d), in the final stage, the as, bs, and cs are estimated while the $\hat{\theta}$s are held fixed. This iterative procedure continues until a convergence criterion is reached. LOGIST cannot provide parameter estimates for examinees with 0 or perfect scores, or for items that all examinees or no examinees answered correctly. This is true for estimation with any of the three models. (See the users manual or Wingersky in Hambleton, 1983, Chapter 3 for more detail on how LOGIST works.)

Accuracy of parameter estimation depends on a number of factors, including the model chosen, the number of item parameters to be estimated (one, two, or three times the number of items), the dimensionality of the data, and the number of items and examinees included in the data set. According to Wingerskey (cited in Hambleton, 1983, Chapter 3), large numbers of examinees and items (on the order of 1000 or more examinees and 40 or more items) should be used when running LOGIST.

Table 2 shows the results for the 14 items in Table 1 obtained by applying LOGIST to the full data set from which the items in Table 1 were drawn. Table 2 (3-parameter model) shows the results from estimating using the 3-parameter model, Table 2 (2-parameter model) shows the results with the *c*s fixed at zero, and Table 2 (1-parameter model) shows the results with the *c*s fixed at zero and the *a* values constrained to be equal. Note that the values of the parameters are not directly comparable. Look at the values for item 11; these values were used to plot the three item characteristic curves in Figure 1.

A more recent estimation program, which is gaining in popularity, is BILOG (Mislevy & Bock, 1984). Like LOGIST, BILOG can handle 1-, 2-, or 3-parameter logistic model estimation. BILOG uses a marginal maximum likelihood solution instead of a joint maximum likelihood solution. Another estimation program, BICAL, was developed solely for use with the Rasch model and is the most popular 1-parameter estimation program (see Wright & Stone, 1979). Other estimation programs also exist (see Hambleton, 1988, pp. 171–172).

TABLE 2

Parameter Values for the 3-, 2-, and 1-Parameter Model, Estimated By LOGIST

Item	a	b	c
	3-parameter model		
1	0.57406	−3.89999	0.20000
2	0.98104	0.05357	0.20239
3	0.80290	0.53538	0.20170
4	2.00000	1.02488	0.13925
5	1.88064	0.24104	0.06373
6	0.83238	−0.94186	0.20000
7	1.01489	0.30096	0.11307
8	1.40447	0.70978	0.11668
9	2.00000	0.39435	0.25016
10	1.71870	1.20411	0.05735
11	1.24899	1.20410	0.13485
12	1.20052	−0.11295	0.23493
13	0.65106	0.95981	0.10245
14	0.78457	0.02969	0.24493
	2-parameter model		
1	1.62047	−1.40369	0.00000
2	0.76428	−0.29026	0.00000
3	0.57242	0.10884	0.00000
4	0.71455	0.97878	0.00000
5	1.45481	0.14637	0.00000
6	0.63411	−1.26397	0.00000
7	0.83281	0.13320	0.00000
8	0.83815	0.57449	0.00000
9	0.84634	−0.05871	0.00000
10	0.90201	1.35452	0.00000
11	0.57959	1.18111	0.00000
12	0.86585	−0.48998	0.00000
13	0.52942	0.80035	0.00000
14	0.62104	−0.44357	0.00000
	1-parameter model		
1	0.77511	−1.99799	0.00000
2	0.77511	−0.28197	0.00000
3	0.77511	0.08073	0.00000
4	0.77511	0.93325	0.00000
5	0.77511	0.25413	0.00000
6	0.77511	−1.11318	0.00000
7	0.77511	0.15080	0.00000
8	0.77511	0.61148	0.00000
9	0.77511	−0.04868	0.00000
10	0.77511	1.47616	0.00000
11	0.77511	0.97321	0.00000
12	0.77511	−0.50946	0.00000
13	0.77511	0.61148	0.00000
14	0.77511	−0.38995	0.00000

Comparing and Evaluating the Models

All three of the logistic models discussed in this module assume a single undimensional trait underlies the data, an assumption seldom, if ever, met in realistic testing situations.

The 1-, 2-, and 3-parameter models differ, however, in the number of parameters they allow to vary. The results obtained, therefore, differ when different models are applied to the same data set. When specifying the model to use, one should select the most stringent model that accurately represents the observed data. Trade-offs will be employed; for example, one might chose to use the 1-parameter model instead of the 3-parameter model with multiple-choice items because of a belief that the *a* and *c* parameters are inestimable or because of the small size of the data set. (LOGIST requires a minimum of approximately 1,000 examinees, BICAL about 200; see Lord cited in Weiss, 1983, Chapter 3.)

Several statistical methods of assessing model fit exist (see, e.g., Hambleton, 1983, 1988; Hambleton & Swaminathan, 1985; Weiss, 1983). One way to examine closeness of fit is to plot the empirical item characteristic curve along with the item characteristic curve based on the parameters estimated by the model. In Figure 6, the empirical curve shows the proportion of examinees at a given θ level who correctly responded to an item (item 11). (As only 3,000 examinees were used, the examinees were grouped in clusters by θ levels.) This empirical curve is based on the θs estimated using the item data for the full data set from which Table 1 was extracted. The item characteristic curve plotted as *fitted model* is based on the parameters estimated by LOGIST and Equation 1. To illustrate the difference between the two ICCs, at $\theta = 1.0$, the 3-parameter model estimated over 60% of the examinees would respond correctly to this particular item; in the empirical data set, about 45% responded correctly. (As the empirical values plotted are based on grouped data, the curve is less smooth.)

Whether the two curves plotted in Figure 6 are "close enough" depends on the proposed uses of the data, on one's views of robustness, and on issues such as the size of the data set used (smaller samples have less precision). For example, establishing an equating chain for a national test may require more accuracy than analyzing a classroom test.

The ICCs in Figure 6, which are based on LOGIST estimates, can be compared with the ICC in Figure 5, which is

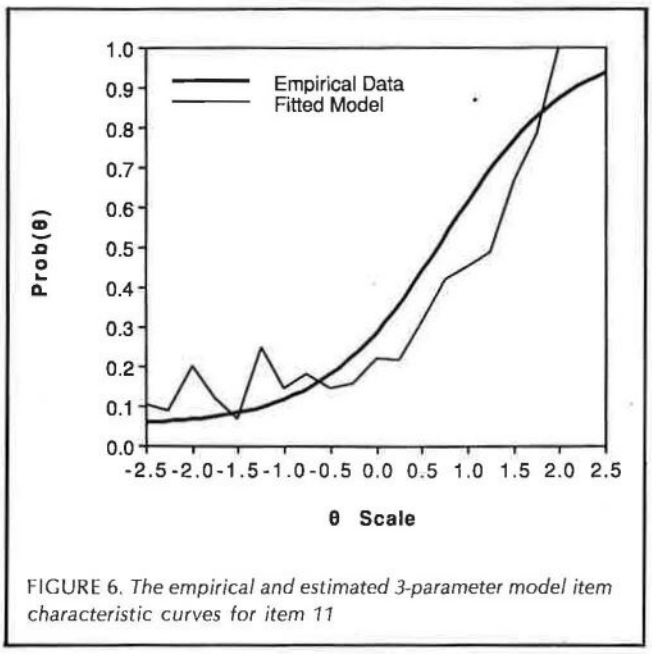


FIGURE 6. *The empirical and estimated 3-parameter model item characteristic curves for item 11*

based on the generated data. To compare actual parameter values, the values would need to be placed on the same scale. Note that it is conceivable that item characteristic curves may diverge in, for example, the tails, but be sufficiently similar in the range of θ values of interest, thus still be useful. Recall, however, that in practice, parameter values will not be available, and only the parameter estimates and empirical data comparisons will be possible. More sophisticated methods for examining model fit exist, but to date no method has been found to be entirely satisfactory.

If one is not satisfied with the data model fit, a different model can be employed or particular examinees or items that clearly do not fit may be deleted (to improve the fit). In some instances, such as when an examinee takes a test at an inappropriate level, this latter option may be desirable; in other instances, philosophical questions, such as whether the statistical model should dictate which items are included on a test, may arise.

Summary

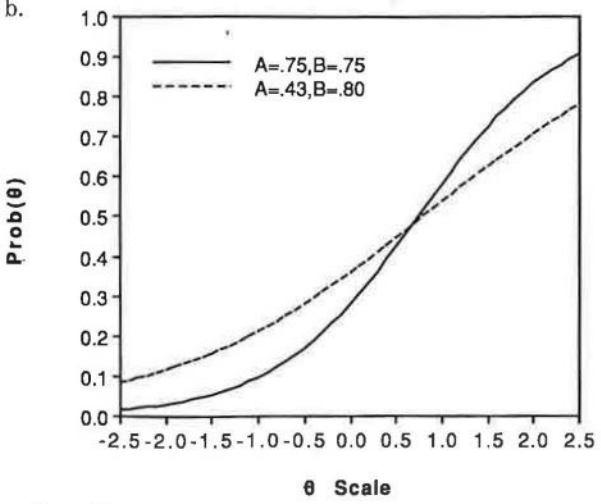
The 3-, 2-, and 1-parameter models were presented and compared. Parameter estimation and model fit were discussed briefly. Before using a model, one needs to evaluate whether the 1-, 2-, or 3-parameter model, all three models, or none of the models adequately fit the data.

Self-Test

1. a. Compute $P(\theta)$ for the following two 2-parameter items for $\theta = -2, -1, 0, +1, +2$.
   Item I. a = .75 b = .75
   Item II. a = .43 b = .80
   b. Sketch the item characteristic curve for each of the items.
   c. Which item is most discriminating for $\theta = 0$?
   d. Which item is most difficult for $\theta = -1$?
   e. Considering only the five θ values listed above, the response pattern 01 would most likely come from an examinee with which θ value?
2. Is the following statement true: "As the 3-parameter model is more general than the 2- or 1- parameter models, it should always be the model of choice." Explain why or why not?
3. Which of the following items would best distinguish between examinees with θ values above and below .66? (Assume all other factors such as content appropriateness, were held equivalent.) Why?
   I. a = 1.12 b = .65 c = .12
   II. a = .80 b = .67 c = .12
4. The 1-parameter model assumes all items within a test have equal discrimination. When is it "sensible" to consider using this model instead of 2- or 3-parameter models?

Answers to Self-Test

1. a. For item I, $P(\theta)$ = .03, .10, .28, .58, and .83; for Item II, $P(\theta)$ = .11, .21, .36, .54, and .71.
   b.



   c. Item I
   d. Item I
   e. The probability of incorrectly answering an item is 1 − $P(\theta)$. Therefore, the probability of the item response vector 01 is (1 − $P(\theta)$) for item I multiplied by $P(\theta)$ for item II. This becomes approximately .11, .19, .26, .23, and .12 for the five values, respectively. It is seen that the value with the highest probability of an item response vector of 01 is $\theta = 0$. This exercise is a much simplified illustration of obtaining a $\hat{\theta}$. In practice, an iterative procedure is used to obtain $\hat{\theta}$s. (See, for example, Baker, 1985.)
2. No; for example, because estimating too many parameters may lessen the accuracy of the parameter estimates.
3. Item I is preferable because it is more highly discriminating in the range of interest.
4. When the data set is too small to use a 2- or 3-parameter model; when computer programs for estimating the 2- and 3-parameter models are not available; when the test has been constructed to meet the 1-parameter assumptions.

Annotated References

Andrich, D. (1988). *Rasch models for measurement.* Beverly Hills, CA: Sage.

An introduction to Rasch modeling for social science measurement, including a comparison with Guttman and Thurstone scaling.

Baker, F. B. (1985). *The basics of item response theory.* Portsmouth, NH: Heinemann Educational Books.

Accompanied by an IBM or Apple disk. Covers models, estimation, and fit; minimal math requirement; user friendly exercises, including graphing the item characteristic curves of 1-, 2-, and 3 parameter models. A good hands-on introduction to IRT.

Hambleton, R. K. (Ed.). (1983). *Applications of item response theory.* Vancouver: Educational Research Institute of British Columbia.

Includes chapters on parameter estimation, choosing a model, model-data fit, and LOGIST.

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

A good overview of IRT, requiring some mathematical background. Includes chapters on estimating parameters and on model-data fit.

Hambleton, R. K. (1988). Principles and selected applications of item response theory. In R. L. Linn (Ed.) *Educational measurement,* (3rd ed., pp. 147–200). New York: Macmillan.

Well-written introduction to IRT, including discussions of assumptions, different IRT models, testing model fit, and applications.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Somewhat mathematical and theoretical, but readable. Discusses item characteristic curves, models, and parameter estimation.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Chapters 16–20 present a mathematically sophisticated treatment of IRT, including Allan Birnbaum's introduction of logistic response models.

Mislevy, R. J., & Bock, R. D. (1984). *BILOG I* (Version 2.2). Mooresville, IN: Scientific Software.

Users manual for BILOG, a marginal maximum likelihood logistic model estimation program.

Warm, T. A. (1978). *A primer of item response theory* (Tech. Rep. No. 941078). Oklahoma City, OK: U.S. Coast Guard Institute.

An introductory text that utilizes a nonmathematical approach to the fundamentals of IRT; quite readable for a novice.

Weiss, D. J. (Ed.). (1983). *New horizons in testing: Latent trait theory and computerized adaptive testing.* New York: Academic Press.

Sections on parameter estimating, including robustness and person-to-model fit. Mathematical demands vary.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST 5* (Version 1.0). Princeton, NJ: Educational Testing Service.

A joint maximum likelihood logistic model estimation program.

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

Concentrates on the Rasch model; includes example of item calibration by hand, model fit, and discussion of BICAL.

Teaching Aids Are Available

A set of teaching aids, designed by Deborah Harris to complement her ITEMS module, "Comparison of 1-, 2-, and 3-Parameter IRT Models," is available at cost from NCME. These teaching aids contain a set of item responses for 200 people to 14 items, along with the a, b, and c parameters and the θ values. As long as they are available, they can be obtained by sending $2.00 to: **Teaching Aids, ITEMS Module #7, NCME, 1230 17th St., NW, Washington, DC 20036.**

Instructional Topics in Educational Measurement Series (ITEMS)

About ITEMS

The purpose of the Instructional Topics in Educational Measurement Series (ITEMS) is to improve the understanding of educational measurement principles. These materials are designed for use by college faculty and students as well as by workshop leaders and participants.

This series is the outcome of an NCME Task Force established in 1985 in response to a perceived need for materials to improve the communication and understanding of educational measurement principles. The committee is chaired by Al Oosterhof, Florida State University. Other members of the committee are Fred Brown, Iowa State University; Jason Millman, Cornell University; and Barbara S. Plake, University of Nebraska.

Topics for the series were identified from the results of a survey of a random sample of NCME members. Authors were selected from persons either responding to a call for authors that appeared in *Educational Measurement: Issues and Practice* or through individual contacts by the committee members. Currently, 17 authors are involved in developing modules. *EM* was selected as the dissemination vehicle for the ITEMS modules. Modules will appear, in a serial fashion, in future issues of *EM*. Barbara S. Plake is serving as editor of the series.

Each instructional unit consists of two parts, (1) instructional module and (2) teaching aids. The instructional modules, which will appear in *EM*, are designed to be learner-oriented. Each module consists of an abstract, tutorial content, a set of exercises including a self-test, and annotated references. The instructional modules are designed to be homogeneous in structure and length. The teaching aids, available at cost from NCME, are designed to complement the instructional modules in teaching and/or workshop settings. These aids will consist of tips for teaching, figures or masters from which instructors can produce transparencies, group demonstrations, additional annotated references, and/ or test items supplementing those included within the learner's instructional unit. The instructional module and teaching aids for an instructional unit are developed by the same author.

To maximize the availability and usefulness of the ITEMS materials, permission is hereby granted to make multiple photocopies of ITEMS materials for instructional purposes. The publication format of ITEMS in *EM* was specifically chosen with ease of photocopying in mind, as the modules appear in consecutive, text-dedicated pages.

The expectation of the Task Force, the editors of *EM,* and NCME is that these modules will be useful in a variety of educational settings. In cooperation with the authors in the series and ad hoc reviewers for the series, this team brings this new series to the NCME readership. If the efforts and enthusiasm of the persons involved in developing, reviewing, and publishing the ITEMS materials is an indication, the series should make a vital contribution to the educational measurement training literature.