DOCUMENT RESUME

ED 250 399 TM 840 754

AUTHOR Reckase, Mark D.; McKinley, Robert L.
TITLE Multidimensional Difficulty as a Direction and a Distance.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE Jun 84
CONTRACT N00014-81-K0817
NOTE 12p.; Paper presented at the Annual Meeting of the Psychometric Society (Santa Barbara, CA, June 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Difficulty Level; Item Analysis; Latent Trait Theory; *Mathematical Models; *Multidimensional Scaling; *Test Items; Test Theory

ABSTRACT

The purpose of this paper is to present a generalization of the concept of item difficulty to test items that measure more than one dimension. Three common definitions of item difficulty were considered: the proportion of correct responses for a group of individuals; the probability of a correct response to an item for a specific person; and the location of the item along a difficulty continuum. This paper defines the difficulty of an item that measures more than one dimension as the direction from the origin of the multidimensional space to the point of greatest discriminating power and the distance from the origin to that point. The direction can be given in terms of angles with the coordinate axes or the corresponding direction cosines. The distance is a signed number using the same units as the coordinate axes. For the unidimensional case, the definition simplifies to the b-parameter from unidimensional item response theory. (BW)

TM

Multidimensional Difficulty as a Direction and a Distance

Mark D. Reckase & Robert L. McKinley
American College Testing Program

The purpose of this paper is to present a generalization of the concept of item difficulty to test items that measure more than one dimension. In developing the generalization of item difficulty, three common definitions of difficulty were considered. The first definition is the proportion of correct responses for a group of individuals. This is the common p-value discussed in many measurement books. This conception of item difficulty yields a result that is specific to the group being used to determine the p-value. It is descriptive of the interaction of the group of persons with the item and does not tell how difficult the item is for any particular person.

A second definition of item difficulty is the probability of a correct response to an item for a specific person. This indication of item difficulty can be determined using an IRT model and an estimate of a person's ability. Of course, for this estimate of item difficulty to be accurate, the IRT model selected must be an accurate representation of the interaction of a person and an item. Unlike the previous definition, this indication of item difficulty is not group specific. The p-value for a specific group can be determined from the probability of a correct response for each person by averaging over persons.

The third definition of item difficulty is the location of the item along a difficulty continuum. The first two definitions yield a value that can be interpreted in this way, but they have the disadvantage of being specific to a group or to a person and not being solely a characteristic of the test item. In IRT, each item is assumed to have a difficulty parameter that is solely a characteristic of the item and is independent of the persons taking the item. The other two types of difficulty statistics can be derived from the IRT model and information about the ability of persons in a group.

Since the first two conceptions of item difficulty can be used as measures of difficulty on a continuum, and since the third definition can be used to derive the statistics used for the previous two, the third definition has been selected as the basis of the generalization of the difficulty concept to more than one dimension. More specifically, the IRT notion of item difficulty will be extended to handle more than one dimension.

Paper presented at the meeting of the Psychometric Society, Santa Barbara, CA, June 1984. This research was supported by Contract Number N00014-81-K0817 from the Personnel and Training Research Programs of the Office of Naval Research.

The Item Response Theory Definition of Item Difficulty

For the unidimensional case, item difficulty is defined in IRT as the point on the ability scale corresponding to the point of inflection of the item characteristic curve (ICC). This point can be determined by solving for the point of inflection of the ICC by taking the second derivative of the item response function, setting it equal to zero, and solving for the non-degenerate root. For example, the difficulty parameter for the two parameter logistic model, given by

$$P(x|\theta, a, b) = \frac{e^{a(\theta - b)}}{1 + e^{a(\theta - b)}} \qquad (1)$$

where θ is the ability parameter and $\underline{a}$ and $\underline{b}$ are item parameters, can be determined by setting the second derivative with respect to the ability parameter equal to zero,

$$\frac{\delta^2 P(x|\theta, a, b)}{\delta\theta^2} = a^2 PQ(1 - 2P) = 0 , \qquad (2)$$

and then solving for its roots. In this case $\underline{a} = 0$, $\underline{P} = 0$, $\underline{Q} = 0$ are degenerate cases, and only $\underline{P} = .5$ yields a meaningful solution. It can be shown that $\underline{P} = .5$ when $\theta = \underline{b}$. Thus, the difficulty of the item is indicated by the point on the θ-scale equal to the $\underline{b}$-parameter in the model. Therefore, $\underline{b}$ is the difficulty parameter for this model.

For unidimensional IRT models, the point on the ability scale indicated by the $\underline{b}$-parameter is the point where the ICC is the steepest. For the two-parameter logistic model, this also indicates the point on the ability scale where the item is most informative.

In order to generalize the IRT definition of item difficulty to multidimensional items, the form of the item response function must first be determined. For the multidimensional case, the item response function gives the probability of a correct response to an item for a person with a particular vector of abilities. A number of different forms have been proposed for this function (see Reckase & McKinley, 1983 for several examples), [illegible] in all cases the functions have been assumed to increase monotonically for all combinations of dimensions. The surface defined by a multidimensional item response function has been labelled an item response surface (IRS).

This paper will use a specific IRS, a multidimensional extension of the two-parameter logistic model (M2PL), to demonstrate the concepts being developed. This model is given by

$$P(x|\theta_j, a_i, d_i) = \frac{e^{x(d_i + a_i' \theta_j)}}{1 + e^{(d_i + a_i' \theta_j)}} \qquad (3)$$

where $P(x_{ij}|\theta_j, a_i, d_i)$ is the probability of response $\underline{x}$ (0 or 1) to Item i for Person j, θ_j is a vector of abilities for Person j, a_i is an item parameter vector for Item i and $\underline{d}_i$ is a scalar item parameter for Item i. The roles of the a_i and $\underline{d}_i$ parameters for this model will be described later in this paper. Two examples of an IRS defined by the M2PL model for the two dimensional case are given in Figure 1.

Figure 1

Examples of Item Response Surfaces
for the M2PL Model

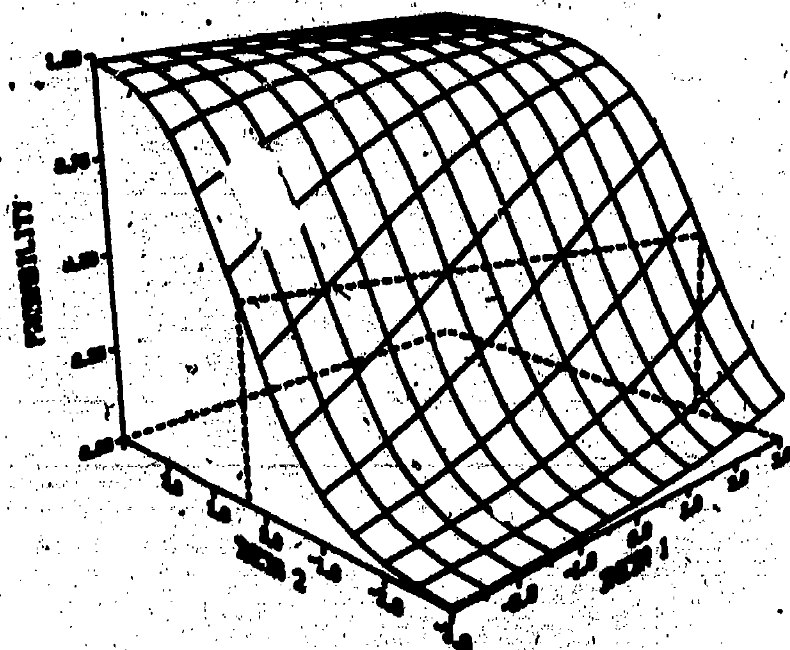

Figure 1a

Parameters

$a_1 = .5$

$a_2 = 1.5$

$d = 1.0$

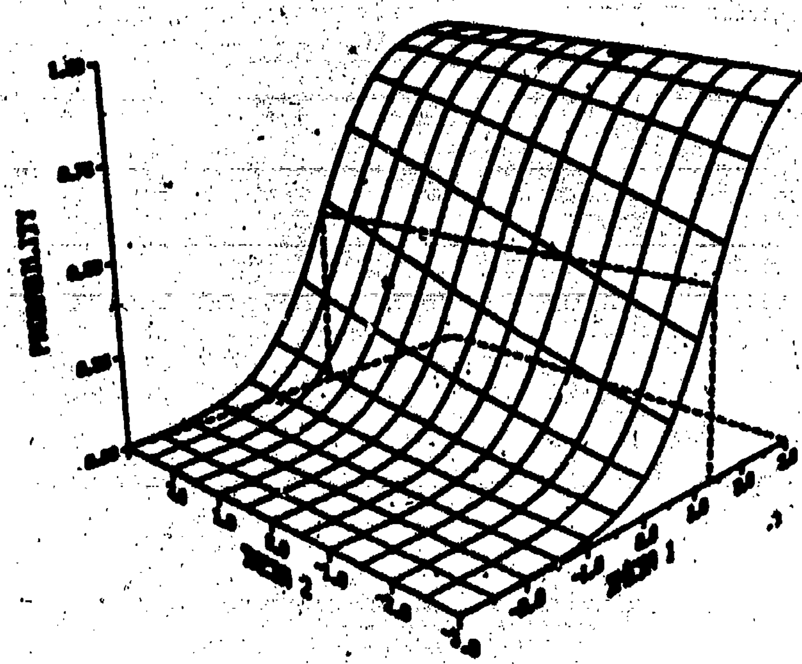

Figure 1b
Parameters

$a_1 = 1.75$

$a_2 = .25$

$d = -1.5$

Definition of Multidimensional Difficulty

The definition of difficulty for multidimensional test items that is proposed in this paper has three purposes: (a) it describes the characteristics of the item so that it can be compared to other items; (b) it gives an indication of the location of the item in the multidimensional ability space; and (c) it tells where the item is most informative.

The definition of difficulty is developed as an extension of the unidimensional IRT definition of difficulty. As given above, difficulty is defined as the location of the point of inflection of the ICC on the ability continuum. For the multidimensional case, difficulty will be defined as the distance and direction of the point of steepest slope from the origin of the multidimensional space. For the M2PL model given previously, this point is the closest point to the origin in the locus of points of inflection for the IRS. While most one-dimensional IRT models have only one point of inflection, the multidimensional models have many. In fact, the locus of points of inflection is usually a function of the ability vector in the multidimensional space. The horizontal lines on the IRSs in Figure 1 shows the locus of points of inflection for these items.

In order to determine the location of a multidimensional item in the multidimensional space, two steps must be performed. First, the locus of points of inflection must be determined. Second, the distance of the locus of points of inflection from the origin of the ability space must be determined. The distance is taken from the origin to the closest point on the locus of points of inflection. Thus, the multidimensional difficulty of a test item has two components: (a) the distance of the locus of points of inflection from the origin of the ability space, and (b) the direction from the origin to the closest point.

Locus of Points of Inflection

For the unidimensional IRT models, the point of inflection of the ICC is determined by taking the second derivative of the item response function and solving for its root. For the multidimensional case, the same procedure is followed, but since the characteristics of the surface depend on the direction, the second directional derivative is used instead of the simple second derivative.

The second directional derivative gives the rate of change of the slope of the surface in a particular direction. For a multidimensional IRT model, the second directional derivative is given by

$$\nabla_{\underline{\phi}}^2 P = \frac{\delta^2 P}{\delta\theta_1^2}\cos^2\phi_1 + \frac{\delta^2 P}{\delta\theta_1 \delta\theta_2}\cos\phi_1 \cos\phi_2 + \dots + \frac{\delta^2 P}{\delta\theta_i \delta\theta_j}\cos\phi_j \cos\phi_i + \frac{\delta^2 P}{\delta\theta_m^2}\cos^2 m , \quad (4)$$

where $\underline{P}$ is the item response function ϕ is the vector of angles with the $\underline{m}$-coordinate axes, and $\theta_1, \dots, \theta_m$ are the m ability dimensions.

For the model presented in Equation 3, the second directional derivative simplifies to

$$\nabla_{\underline{\phi}}^2 P = PQ(1 - 2P)\,(a_1 \cos\phi_1 + a_2 \cos\phi_2 + \ldots + a_m \cos\phi_m)^2 , \quad (5)$$

where all of the variables have been defined previously.

As with the 2-parameter logistic model, the only non-degenerate solution is the case when $\underline{P} = .5$. When $\underline{P} = .5$, the exponent of Equation 1 must be 0. Therefore, the locus of points of inflection for this model is given by

$$d_i + a_i' \theta_j = 0 \quad (6)$$

This is the equation for a hyperplane in a $\underline{m}$-dimensional space.

The proposed indicators of difficulty for a multidimensional test item are the shortest distance between the locus of points of inflection and the origin of the space, and the direction used to obtain the shortest distance. The shortest distance between the origin and the hyperplane of inflection is along a perpendicular to the hyperplane. The direction cosines of a line that is perpendicular to the hyperplane of inflection are given by

$$\cos\phi_j = \frac{a_{ji}}{\sqrt{\Sigma a_{ji}^2}} \quad (7)$$

where ϕ_j is the angle between the line and the ith dimensional axis, and $\underline{a}_{ji}$ is the jth term in the item parameter vector a_i.

The distance of the hyperplane of inflection from the origin for this model is given by

$$D_i = \frac{-d_i}{\sqrt{\Sigma a_{ji}^2}} \quad (8)$$

where $\underline{d}_i$ is the scalar parameter in the exponent of the model and $\underline{a}_{ji}$ has been previously defined.

The values given by Equations 7 and 8 have a fairly clear interpretation. The angles defined by Equation 7 through the direction cosines indicate the direction with respect to the ability dimensions in which the item provides the most information. Since the direction specified is perpendicular to the line of inflection, the slope of the IRS is steepest in that direction. Thus the item is best at measuring a weighted composite of abilities defined by

$$\text{Composite} = \sum_{i=1}^{m} \cos\phi_i \, \theta_i . \tag{9}$$

If a set of items were selected that have the same direction cosines, the test would operate as if it were unidimensional since all of the items discriminate best for the same composite of abilities. If one of the direction cosines were 1.0, the rest would be 0.0, and the item would be a pure measure of one of the abilities.

The distance of the line of inflection from the origin can be interpreted in much the same way as the b-parameter for unidimensional IRT models. For two items with the same direction cosines, the item with the larger positive D-value will have a smaller proportion of individuals obtaining a correct response. Negative values of D generally indicate easy items, but only items with the same direction cosines can be directly compared.

An Illustrative Example

In order to demonstrate the application of the concept of multidimensional item difficulty, a 30-item multiple choice examination with two distinctly different kinds of items, spelling and grammar, was analyzed using the M2PL model using a program developed by McKinley & Reckase (1983). The data on this test were obtained from the administration of the items to 1000 students at the University of Texas at Austin as part of an entrance battery.

Table 1 gives the item parameter estimates for a two dimensional solution for the M2PL model as well as the direction relative to the θ_1-axis and the distance to the point of inflection closest to the origin. Note that the angles with the θ_1-axis for the first 15 items tend to cluster near 0° ($\bar{\phi} = 11.09$), indicating that these items are best at measuring the θ_1 ability. Since these items are all punctuation items, θ_1 can be labelled as a punctuation dimension. The angles with the θ_1-axis of the second 15 items tend to cluster around 90° ($\bar{\phi} = 77.15$). This implies that these items are best at measuring θ_2. Since these items are all grammar items, θ_2 can be labelled as a grammar dimension.

Table 1

Item Parameter Estimates and Difficulty Statistics For 30 Punctuation and Grammar Items

Item	Item Parameter Estimates			Directional Estimates	
	d	a_1	a_2	ϕ	D
1	3.17	1.42	0.12	4.83	2.22
2	1.45	1.36	0.20	8.37	1.05
3	0.66	0.79	0.28	19.52	0.79
4	0.22	0.93	0.20	12.14	0.23
5	0.97	0.66	0.17	14.44	1.42
6	1.40	0.78	0.16	11.59	1.76
7	1.59	0.97	0.21	12.22	1.60
8	1.80	1.26	0.17	7.68	1.42
9	0.43	0.65	0.17	14.66	0.64
10	2.05	1.00	0.16	9.09	2.02
11	0.80	1.11	0.14	7.19	0.72
12	3.22	1.12	-0.08	4.09	2.87
13	-0.94	0.69	0.41	30.72	-1.17
14	2.27	1.39	0.23	9.40	1.61
15	2.04	1.57	0.01	0.36	1.30
16	1.12	0.47	0.89	62.16	1.11
17	-2.68	0.33	0.85	68.78	-2.94
18	-0.06	0.20	0.41	64.00	-0.13
19	1.44	0.02	0.57	87.99	2.52
20	0.34	-0.22	0.51	113.33	0.61
21	1.58	0.10	0.73	82.20	2.14
22	-0.30	0.09	0.78	83.42	-0.38
23	0.50	0.31	0.55	60.59	0.79
24	-0.85	0.26	1.28	78.52	-0.65
25	-0.30	-0.26	0.61	113.09	-0.45
26	1.04	0.22	0.32	55.49	2.68
27	-0.79	0.25	0.73	71.10	-1.02
28	0.40	0.20	0.47	66.95	0.78
29	-1.66	0.07	0.80	85.00	-2.07
30	1.11	0.26	0.55	64.70	1.82

The D-values in Table 1 indicate the distance of the nearest point of inflection to the origin of the θ-space. If two items measure effectively in the same direction from the origin, the D-values indicate the relative difficulty of the items. For example, items 20 and 25 are measuring approximately the same combination of abilities. Since the D-value for Item 20 is .61 and the D-value for Item 25 is -.45, Item 20 is estimated to be more difficult than Item 25 on the particular composite measured by the items. When the direction of the item is taken into account, the D-values can be interpreted much like the b-parameter estimates from unidimensional item response theory.

Summary and Conclusions

A definition of item difficulty has been suggested for use with items that require ability on more than one dimension for a correct response. This definition has two components: (a) the direction from the origin of the multidimensional space for which the item provides the most information; and (b) the distance from the origin of the space to the point of steepest slope on the IRS. This definition was demonstrated for the multidimensional extension of the two-parameter logistic model.

The statistics provided by this definition can be directly applied to the process of test construction. If a test that measures the abilities that define the latent space is desired, items should be selected that have directions that parallel the axes of the space. However, if it is merely desirable to construct a test that operates as if it were unidimensional, then items that have the same direction should be selected. These items measure the same composite of abilities.

The D-statistics provided by the definition gives the distance from the origin to the nearest point of inflection. This value can be interpreted in the same way as the b-parameter from unidimensional IRT model when items measure in the same direction.

References

McKinley, Robert L. and Reckase, Mark D. MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. Behavior Research Methods and Instrumentation, 1983, 15(3), 389-390.

Reckase, Mark D. and McKinley, Robert L. The definition of difficulty and discrimination for multidimensional item response theory models. Paper presented at the meeting of the American Educational Research Association, Montreal, April 1983.

Abstract

Multidimensional difficulty as a direction and a distance

The paper defines the difficulty of an item that measures more than one dimension as the direction from the origin of the multidimensional space to the point of greatest discriminating power and the distance from the origin to that point. The direction can be given in terms of angles with the coordinate axes or the corresponding direction cosines. The distance is a signed number using the same units as the coordinate axes. For the unidimensional case, the definition simplifies to the b-parameter from unidimensional item response theory.