The Difficulty of Test Items That Measure More Than One Ability

Mark D. Reckase
The American College Testing Program

Many test items require more than one ability to obtain a correct response. This article proposes a multidimensional index of item difficulty that can be used with items of this type. The proposed index describes multidimensional item difficulty as the direction in the multidimensional space in which the item provides the most information and the distance in that direction to the most informative point. The multidimensional difficulty is derived for a particular item response theory model and an example of its application is given using the ACT Mathematics Usage Test.

Statistics that describe the characteristics of test items are routinely used to assist in the test construction process. These statistics are often used to help produce equivalent forms of tests and to produce tests according to detailed specifications. Sometimes they are used to select items that will yield test forms that measure a single trait or dimension.

Most, if not all, of the statistics that are commonly used to describe a test item assume that the item measures a single trait or dimension. The unidimensional item response theory (IRT) procedures (see Traub & Lam, 1985, for a summary of these procedures) make the assumption of a single trait directly, but even the traditional statistics, such as p and r_{bis}, make an implicit assumption that items can be ordered in difficulty along a single continuum and that a single dimension is being measured by the test.

Yet most items are multidimensional in some sense, and, depending on the strength of the multiple dimensions, these unidimensional statistics may not be appropriate. Some items measure a fairly strong first dimension with only minor higher-order dimensions. Vocabulary and some mathematics computation items may fall into this category. Although there may be many different types of vocabulary words and synonyms, one common dimension seems to be measured by vocabulary tests. For these types of items, unidimensional statistics seem reasonable.

Some items clearly require more than one distinct ability to arrive at a correct response. Mathematical "story problems" are the most common example of this type of item; both mathematical and verbal skills are required to obtain a correct answer. Current measures of item characteristics are inappropriate for this type of item. For example, if a proportion-correct difficulty index is used to describe items of this type, the ranking of items on difficulty may be quite different for an examinee sample that differs mainly on verbal skill than for an examinee sample that differs mainly on mathematical skill.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 9, No. 4, December 1985, pp. 401–412

0146-6216/85/040401-12$1.85

What is needed is a means of describing the characteristics of an item that takes into account the dimensionality of the skills required to solve the item. Such a statistic can then be used to determine how or if it is possible to compare items that measure different combinations of abilities. The statistic can be used in test construction to ensure that a test matches predefined characteristics and, if desirable, to form tests that measure a single characteristic.

The goal of this article is to describe a means of determining the difficulty of an item that measures more than one dimension. This multidimensional measure of difficulty is based on a multidimensional generalization of IRT concepts. To demonstrate the usefulness of this concept, it is compared to several commonly used unidimensional item statistics.

Method

The proposed definition of a measure of multidimensional item difficulty (MID) is based on three general assumptions. First, it is assumed that the probability of answering an item correctly increases monotonically with an increase in each dimension being measured. Second, it is assumed that it is desirable to locate an item at a single point in a multidimensional space. Previous work (Reckase & McKinley, 1983) defined MID as the locus of points of inflection of a multidimensional item response surface (IRS). That definition resulted in the MID for a multidimensional item being characterized by a hypersurface in an n-dimensional space. Using that definition, it was very difficult to determine whether two items were measuring the same combination of traits and the comparison of the difficulty of the items was unwieldy. Locating each item at a point greatly simplifies the use of the MID concept.

The third assumption is that the most reasonable point to use in defining the MID for an item in the multidimensional space is the point where the item is most discriminating. This is the point where the item provides the most information about the person being measured.

On the basis of these three assumptions, a general technique for determining the MID will be specified. However, in order to demonstrate the procedure, a particular model is needed. For the purpose of demonstration, the multidimensional extension of the two-parameter logistic (M2PL) model has been selected (McKinley & Reckase, 1983a). This model has been selected because it meets the first assumption and because estimation procedures are available for the parameters (McKinley & Reckase, 1983b).

The M2PL model can be described by the following equation:

$$P(x_{ij} = 1|\mathbf{a}_i, d_i, \boldsymbol{\theta}_j) = \frac{e^{(\mathbf{a}_i' \boldsymbol{\theta}_j + d_i)}}{1 + e^{(\mathbf{a}_i' \boldsymbol{\theta}_j + d_i)}} \quad , \tag{1}$$

where $P(x_{ij} = 1|\mathbf{a}_i, d_i, \boldsymbol{\theta}_j)$ is the probability of a correct response to item i by person j;
$\mathbf{a}_i$ is a vector of discrimination parameters;
d_i is a scalar parameter that is related to the difficulty of the item; and
$\boldsymbol{\theta}_j$ is a vector of ability parameters.

The exponent can also be expressed as

$$\sum_{k=1}^{n} a_{ik}(\theta_{jk} - b_{ik}) \quad , \tag{2}$$

where n is the number of dimensions,
a_{ik} is an element of $\mathbf{a}_i$,
θ_{jk} is an element of $\boldsymbol{\theta}_j$, and
$d_i = -\sum_{k=1}^{n} a_{ik} b_{ik}$.

In this form, it is more similar to the usual expression for the two-parameter logistic model.

The M2PL model defines an IRS that is monotonically increasing in probability as the elements of $\boldsymbol{\theta}_j$ increase. An example of the IRS defined by the model for the two-dimensional case is shown in Figure 1. This IRS was generated using $a_{i1} = .75$, $a_{i2} = 1.2$, and $d_i = -1$. This surface clearly meets the first assumption given above.

The goal of the following analysis is to determine the point in the θ space at which the IRS has the maximum slope. However, the slope along a surface differs depending on the direction taken relative to the surface. That is, the slope at a point can be quite different in one direction than in another. In order to develop a definition for MID that is easily usable, the slope at a point in the θ space will always be determined using the direction from the origin of the θ space to the point. This choice of a direction has implications that are described below.

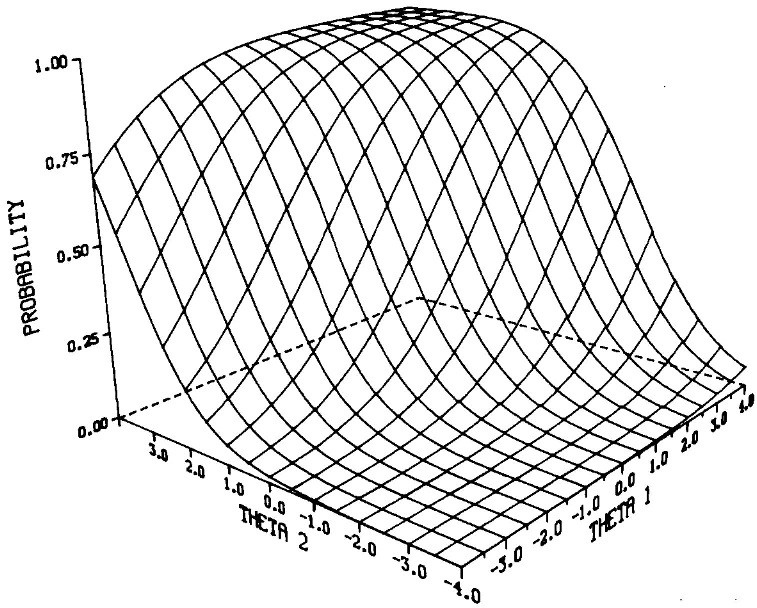
The procedure to be used to find the point of maximum slope from the origin involves two steps. First, the point of maximum slope in a particular direction is determined. Then, the slopes in each direction are analyzed to determine the direction that yields the overall maximum.

In order to facilitate the analysis, the model given in Equation 1 is translated to polar coordinates. That is, each θ_{jk} is replaced by $\theta_j \cos \alpha_{jk}$, where θ_j is the distance from the origin to $\boldsymbol{\theta}_j$ and α_{jk} gives the angle from the kth axis to the point.

The revised expression for the M2PL model is

$$P(x_{ij} = 1|\mathbf{a}_i, d_i, \boldsymbol{\alpha}_j, \theta_j) = \frac{e^{\left(\sum_{k=1}^{n} a_{ik}\theta_j \cos \alpha_{jk} + d_i\right)}}{1 + e^{\left(\sum_{k=1}^{n} a_{ik}\theta_j \cos \alpha_{jk} + d_i\right)}} \quad . \tag{3}$$

Figure 1
Item Response Surface for the M2PL Model
With Parameters $a_1 = .75$, $a_2 = 1.2$, and $d = -1$



In order to find the maximum slope in a particular direction $\boldsymbol{\alpha}_j$, the second derivative is taken with respect to θ_j and solved for its zero point. This analysis gives the point of inflection of the IRS in a particular direction. For the M2PL model, the second derivative with respect to θ_j is given by

$$\frac{\delta^2 P(x_{ij} = 1|\mathbf{a}_i, d_i, \boldsymbol{\alpha}_j, \theta_j)}{\delta\theta_j^2} = \left(\sum_{k=1}^{n} a_{ik}\cos\alpha_{jk}\right)^2 P_{ij}(1 - 3P_{ij} + 2P_{ij}^2) \quad , \tag{4}$$

where $P_{ij} = P(x_{ij} = 1|\mathbf{a}_i, d_i, \boldsymbol{\alpha}_j, \theta_j)$. Equation 4 is equal to zero when $P_{ij} = .5$. Thus, the slope in direction $\boldsymbol{\alpha}_j$ is at its maximum when the IRS crosses the .5 plane.

The slope of the surface in direction $\boldsymbol{\alpha}_j$ is given by the first derivative of the model with respect to θ_j,

$$\frac{\delta P(x_{ij} = 1|\mathbf{a}_i, d_i, \boldsymbol{\alpha}_j, \theta_j)}{\delta\theta_j} = P_{ij}(1 - P_{ij}) \sum_{k=1}^{n} a_{ik}\cos\alpha_{jk} \quad . \tag{5}$$

When $P_{ij} = .5$, the slope is equal to

$$\tfrac{1}{4}\sum_{k=1}^{n} a_{ik}\cos\alpha_{jk} \quad . \tag{6}$$

When $\alpha_{jk} = 0$ and $\alpha_{jl} = 90$ for $l \neq k$, indicating that the slope of the surface at $P_{ij} = .5$ is determined parallel to axis k, $\cos\alpha_{jk} = 1$ and $\cos\alpha_{jl} = 0$. The slope parallel to axis k is then given by $1/4\ a_{ik}$. Thus a_{ik} is related to the slope of the IRS at the point of steepest slope, as is the case for the unidimensional IRT models.

To find the direction of steepest slope, Expression 6 is differentiated with respect to $\cos\alpha_{jk}$ and solved for zero. However, before performing the differentiation, the constraint that $\sum_{k=1}^{n}\cos^2\alpha_{jk} = 1$ is added to the expression for the slope by setting

$$\cos^2\alpha_{jn} = 1 - \sum_{k=1}^{n-1}\cos^2\alpha_{jk} \quad . \tag{7}$$

This constraint is equivalent to assuming that the θ axes are orthogonal. The differentiation results in $n - 1$ equations of the form

$$a_{i1} - a_{in}\left(\frac{\cos\alpha_{j1}}{\cos\alpha_{jn}}\right) = 0 \quad ,$$

$$a_{i2} - a_{in}\left(\frac{\cos\alpha_{j2}}{\cos\alpha_{jn}}\right) = 0 \quad ,$$

$$\vdots$$

$$a_{in-1} - a_{in}\left(\frac{\cos\alpha_{jn-1}}{\cos\alpha_{jn}}\right) = 0 \quad . \tag{8}$$

This system of equations can be solved for $\cos\alpha_{ik}$ in terms of the a parameters. The result is

$$\cos\alpha_{ik} = \frac{a_{ik}}{\left(\sum_{k=1}^{n} a_{ik}^2\right)^{1/2}} \quad . \tag{9}$$

Up to this point, the $\boldsymbol{\alpha}_j$ vector was considered as a person parameter that was used to convert from rectangular to polar coordinates. When the direction that yields the maximum slope is found from the system of equations given in Equation 8, $\boldsymbol{\alpha}$ changes to an item parameter. Therefore, it will be denoted as $\boldsymbol{\alpha}_i$ in the following equations.

The above derivation gives the angle from the origin to the point of steepest slope. To determine the signed distance, D_i, to the point of steepest slope, Equation 9 can be substituted for $\cos \alpha_{ik}$ in Equation 3 and the resulting equation can be solved for θ_j for $P_{ij} = .5$, the value of the probability when the slope is maximum. The result is

$$D_i = \frac{-d_i}{\left(\sum_{k=1}^{n} a_{ik}^2\right)^{1/2}} \quad , \tag{10}$$

where D_i is an item parameter.

Thus, to describe the difficulty of a multidimensional item, it is proposed that a set of statistics be reported: the distance to the point of steepest slope in a direction from the origin, and the angles, or direction cosines, needed to describe the direction.

At this point it may prove helpful to give an example of the MID for an item. The item shown in Figure 1 will be used for this purpose. The a parameters for the item were $a_{i1} = .75$ and $a_{i2} = 1.2$, therefore the direction cosines are $\cos \alpha_{i1} = .53$ and $\cos \alpha_{i2} = .85$ corresponding to angles of 58° and 32° with the θ_1 and θ_2 axes, respectively. The signed distance in the direction to the point of steepest slope is .71. This corresponds to a θ_1 coordinate of .37 and a θ_2 coordinate of .60. The distance can be interpreted much like a b parameter from unidimensional IRT, indicating that the item is fairly difficult for a population centered at the origin.

In order to demonstrate the usefulness of the MID statistics for the analysis of actual test data, the statistics were computed using the item response data from a representative sample of 1,000 students who took the American College Testing (ACT) Assessment Mathematics Usage Test in February 1983. The M2PL parameter estimates for the 40 items on the test were determined using the MAXLOG program (McKinley & Reckase, 1983b). As a basis for comparison, a traditional item analysis and a three-parameter logistic calibration using LOGIST (Wingersky, Barton, & Lord, 1982) were performed on the same data set.

Results

The parameter estimates from the M2PL model, the directions, and the distances for a two-dimensional analysis of the ACT Mathematics Usage Test are presented in Table 1. A two-dimensional analysis was performed to keep the demonstration of the procedures simple. Also, the method for determining the appropriate number of dimensions has not been solved.

The second, third, and fourth columns of Table 1 give the parameter estimates from the M2PL model. Notice that the lower numbered items (the easier items in the sense of proportion-correct difficulty) tend to have high values for a_{i1}, while the higher numbered items (the more difficult items) tend to have high values for a_{i2}. The values of d_i reflect the difficulty of the items.

The next two columns show the direction cosines for each item and the following two columns give the corresponding angles with θ_1 and θ_2 axes. Note that the angles must add to 90° since the solution is in two dimensions. The angles with the axes reflect the pattern present in the a parameter estimates. The easier items tend to be clustered along the θ_1 axis and the more difficult items tend to have directions close to the θ_2 axis.

The last column in the table gives the distance measure for each item. The distances tend to reflect the difficulty ordering of the items.

Table 1
Item Parameters, Directions
and Distances for the Items
in the ACT Assessment Mathematics Usage Test

Item Number	a_{i1}	a_{i2}	d_i	$\cos \alpha_{i1}$	$\cos \alpha_{i2}$	α_{i1}	α_{i2}	D_i
1	1.81	.86	1.46	.90	.43	25	65	-.73
2	1.22	.02	.17	1.00	.02	1	89	-.14
3	1.57	.36	.67	.97	.22	13	77	-.42
4	.71	.53	.44	.80	.60	37	53	-.50
5	.86	.19	.10	.98	.21	12	78	-.11
6	1.72	.18	.44	.99	.10	6	84	-.25
7	1.86	.29	.38	.99	.15	9	81	-.20
8	1.33	.34	.69	.97	.25	14	76	-.50
9	1.19	1.57	.17	.60	.80	53	37	-.09
10	2.00	.00	.38	1.00	.00	0	90	-.19
11	.87	.00	.03	1.00	.00	0	90	-.03
12	2.00	.98	.91	.90	.44	26	64	-.41
13	1.00	.89	-.49	.75	.66	42	48	.37
14	1.22	.14	.54	.99	.11	7	83	-.44
15	1.27	.47	.29	.94	.35	20	70	-.21
16	1.35	1.15	-.21	.76	.65	40	50	.12
17	1.06	.45	.08	.92	.39	23	67	-.07
18	1.92	.00	.12	1.00	.00	0	90	-.06
19	.96	.22	-.30	.97	.22	13	77	.30
20	1.20	.12	-.28	.99	.10	6	84	.23
21	1.41	.04	-.21	.99	.03	2	88	.15
22	1.54	1.79	.02	.65	.76	49	41	-.01
23	.54	.23	-.69	.92	.39	23	67	1.18
24	1.53	.48	-.83	.95	.30	17	73	.52
25	.72	.55	-.56	.79	.61	37	53	.62
26	.51	.65	-.49	.62	.79	52	38	.59
27	1.66	1.72	-.38	.69	.72	46	44	.16
28	.69	.19	-.68	.96	.27	15	75	.95
29	.88	1.12	-.91	.62	.79	52	38	.64
30	.68	1.21	-1.08	.49	.87	61	29	.78
31	.24	1.14	-.95	.21	.98	78	12	.82
32	.51	1.21	-1.00	.39	.92	67	23	.76
33	.76	.59	-.96	.79	.61	38	52	1.00
34	.01	1.94	-1.92	.01	1.00	90	0	.99
35	.39	1.77	-1.57	.22	.98	78	12	.87
36	.76	.99	-1.36	.61	.79	52	38	1.09
37	.49	1.10	-.81	.41	.91	66	24	.67
38	.29	1.10	-.99	.25	.97	75	15	.87
39	.48	1.00	-1.56	.43	.90	64	26	1.41
40	.42	.75	-1.61	.49	.87	61	29	1.87

The results given in Table 1 can be presented graphically as vectors in a two-dimensional space. Such a representation is given in Figure 2. Note that the items with negative values for D_i tend to cluster along the θ_1 axis, while the more difficult items are more scattered.

The traditional difficulty (p value) and discrimination (biserial correlation) statistics for the 40 mathematics items are given in Table 2. The statistics show that the items are arranged in approximate order of difficulty, and that they tend to have uniformly high values of discrimination.

Figure 2
Representation of the ACT Mathematics Usage
Items as a Direction and a Distance

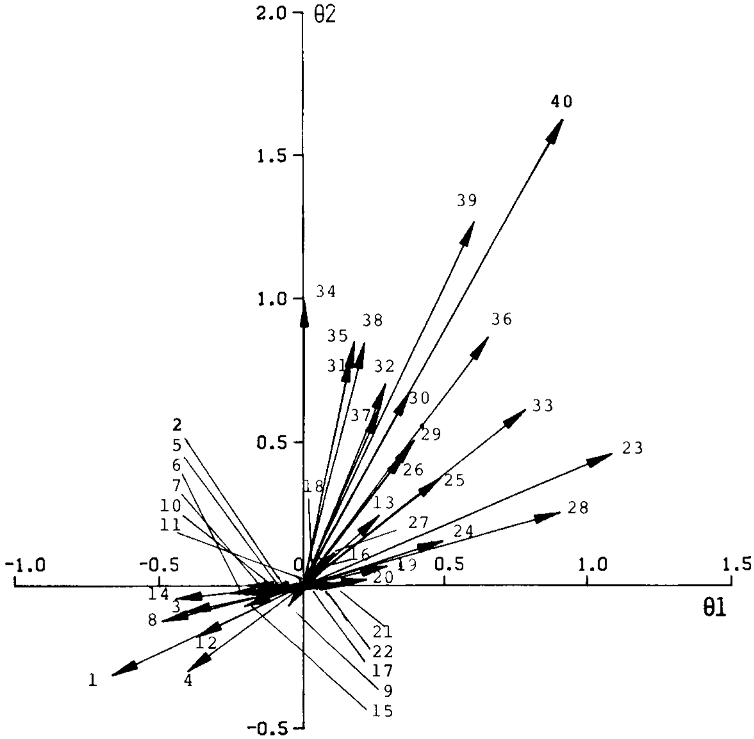


Table 3 presents the parameter estimates for the three-parameter logistic model. The estimates for the parameters were obtained using LOGIST 5 on the sample of 1,000. The *a* parameter estimates given in the table tend to be higher for the more difficult items. The *b* parameter estimates roughly indicate the ordering of the items according to difficulty.

In order to give some basis for comparison of the statistics presented, Pearson product-moment correlations were computed for all possible pairs of statistics. These correlations are only reported as an aid for interpretation of the data. In some cases, nonlinear relationships are present between the coefficients, so the correlations should be interpreted cautiously. The correlations for all of the statistics are presented in Table 4.

The strongest relationships indicated in Table 4 are those among the difficulty indices. The values of *p, d, D,* and *b* are all highly interrelated. All seem to be an indicator of the probability of answering an item correctly for the group of individuals that was tested.

A second relationship that is of interest concerns the measures of discrimination r_{bis} and *a*. These two values only correlated .14. An analysis of the scatterplot of these statistics indicated that the low correlation was the result of several difficult items that had relatively low r_{bis} values and high *a* values (i.e., Number 34). The r_{bis} values for these items may have been deflated by guessing effects.

The direction measure for the items, $\cos \alpha_{i1}$, is most highly related to *p*, indicating that the dimension measured by the items changes with the difficulty of the item. The high correlations with the other M2PL model parameters are artifacts. For example, $\cos \alpha_{i1}$ and $\cos \alpha_{i2}$ are functionally related and must be

Table 2
Proportion Correct and Biserial Correlation
for each Item on the ACT Mathematics Usage Test

Item Number	p	r_{bis}
1	.70	.48
2	.53	.41
3	.60	.48
4	.59	.36
5	.52	.37
6	.56	.49
7	.55	.51
8	.62	.44
9	.50	.49
10	.54	.54
11	.51	.32
12	.60	.53
13	.39	.49
14	.60	.41
15	.54	.47
16	.44	.55
17	.51	.44
18	.51	.49
19	.43	.39
20	.44	.43
21	.46	.45
22	.47	.56
23	.34	.29
24	.35	.53
25	.38	.40
26	.39	.35
27	.41	.60
28	.35	.33
29	.32	.49
30	.30	.45
31	.32	.33
32	.31	.40
33	.30	.42
34	.22	.32
35	.26	.41
36	.25	.47
37	.34	.40
38	.31	.34
39	.22	.39
40	.20	.31

highly correlated. The *a* parameter from the three-parameter logistic model is most highly related to a_2, indicating that LOGIST is estimating ability from the second M2PL dimension.

Discussion

The purpose of this article was to present a definition for multidimensional item difficulty (MID) and to demonstrate its use for a particular set of test items. The definition presented defined MID as the direction

Table 3
Estimated Item Parameters
for the Three-Parameter Logistic Model

Item Number	a	b	c
1	1.08	-.64	.12
2	.66	.08	.12
3	.97	-.25	.12
4	.47	-.29	.12
5	.50	.18	.12
6	.93	-.10	.11
7	1.21	.05	.15
8	.77	-.35	.12
9	1.28	.34	.20
10	.97	-.17	.02
11	.40	.29	.12
12	1.44	-.22	.12
13	1.23	.72	.16
14	.65	-.29	.12
15	.91	.09	.15
16	1.36	.38	.12
17	1.09	.47	.24
18	.85	.03	.06
19	.73	.71	.16
20	.81	.58	.15
21	.84	.44	.13
22	1.46	.30	.14
23	.41	1.61	.12
24	.89	.54	.00
25	.91	1.00	.18
26	.57	.98	.13
27	1.36	.35	.06
28	.93	1.36	.22
29	1.31	.91	.12
30	1.43	1.11	.15
31	1.60	1.44	.23
32	1.34	1.22	.18
33	1.46	1.18	.17
34	2.00	1.55	.15
35	1.73	1.29	.14
36	.79	1.15	.03
37	1.05	1.15	.18
38	1.22	1.45	.20
39	1.25	1.52	.11
40	1.94	1.67	.13

from the origin of the multidimensional space to the point of greatest discrimination for the item and the distance to that point. For the two-dimensional case, two statistics are required to present this information, the angle with one of the axes, and the distance along the vector to the point of maximum discrimination. In n dimensions, n statistics are needed to specify the MID: $n - 1$ angles and a distance. The nth angle can be determined from the other $n - 1$ angles.

An alternative definition of MID could be the rectangular coordinates of the point of greatest discrimination in the direction $\boldsymbol{\alpha}_i$ from the origin of the θ space. The coordinate along dimension k is given by

Table 4
Correlation between Item Statistics
for the Items of the ACT Assessment Mathematics Usage Test

Statistic	Statistic 2	3	4	5	6	7	8	9	10	11	12	13
1. p	.38	.75	-.47	.99	.72	-.69	-.71	.71	-.96	-.49	-.95	-.13
2. r_{bis}		.80	.16	.41	.31	-.18	-.24	.24	-.49	.14	-.55	-.44
3. a_{i1}			-.34	.78	.71	-.66	-.70	.70	-.75	-.21	-.82	-.47
4. a_{i2}				-.46	-.81	.87	.86	-.86	.33	.74	.38	.16
5. d_i					.74	-.67	-.71	.71	-.94	-.47	-.93	-.12
6. $\cos \alpha_{i1}$						-.92	-.97	.97	-.63	-.69	-.69	-.30
7. $\cos \alpha_{i2}$							.98	-.98	.61	.65	.65	.30
8. α_{i1}								-1.00	.63	.69	.68	.31
9. α_{i2}									-.63	-.69	-.68	-.31
10. D_i										.42	.96	.15
11. a											.40	.24
12. b												.34
13. c												

$$\theta_{ik} = D_i \cos \alpha_{ik} = \frac{-d_i a_{ik}}{\sum_{k=1}^{n} a_{ik}^2} . \qquad (11)$$

This definition of MID was not judged to be as useful as the one proposed because it cannot be used as readily to determine whether two items measure the same composite of abilities, and it does not have as clear a connection to the proportion correctly responding to the item. For the proposed definition, two items with the same $\boldsymbol{\alpha}_i$ measure the same composite of abilities regardless of their difficulty. It would not be immediately obvious that two sets of θ coordinates were on the same line through the origin, indicating that the items measured the same composite. The D_i value in the proposed definition is highly related to the proportion of the population responding correctly, and it can be interpreted much as a b value from unidimensional IRT. A set of coordinates does not present this information about the difficulty of the item in as usable a form.

The direction and distance are not unique for an item any more than any other IRT parameters are unique. They are only uniquely defined when the origin and unit of measurement of the complete latent space are specified. If a different origin is used, different units are specified, or if the axes are rotated, the direction and distance will change. Work is currently being done to determine the procedures for translating from one specification of a latent space to another. Such procedures will be needed to link multidimensional calibrations or for equating multidimensional tests.

The MID information for the ACT Mathematics Usage Test revealed some interesting information about the test. The easier items tended to be measures of θ_1, while the more difficult items tended to give more information about θ_2. From an analysis of items that are most highly related to θ_1, it seems that the dimension is related to mathematics problems with a strong verbal component. For example, Item 6 is almost a pure measure of θ_1 and it clearly requires reading comprehension skills. Item 20 is a more difficult item of this same type (see Figure 3 for these items).

Item 34, on the other hand, is almost a pure measure of θ_2. Item 9 is an easier version of this type of item. Both of these items require very little verbal comprehension skill. Both of these items are also presented in Figure 3.

In order to compare the difficulty level of items, the items must measure the same composite of abilities, that is, they must have the same direction. Thus, it is reasonable to compare the difficulties of

Figure 3
Items Measuring Each of the Two Dimensions

θ_1

6. Sheila's salary is $110 per day. Due to financial problems in her company, her employer has asked Sheila to take a 10% cut in pay. How much will Sheila be earning per day if she takes the cut in pay?

F. $ 11
G. $ 99
H. $100
J. $109
K. $121

20. A serving of a certain cereal, with milk, provides 35% of the potassium required daily by the average adult. If a serving of this cereal with milk contains 112 milligrams of potassium, how many milligrams of potassium does the average adult require each day?

F. 35
G. 39
H. 147
J. 320
K. 392

θ_2

9. $|-5| + |6| + (-5) + 6 = ?$

A. −22
B. −10
C. 2
D. 10
E. 12

34. Which line is parallel to $y = 3x + 1$ and intersects $y = 6x - 1$ on the y-axis?

F. $y = 3x - 1$
G. $y = 2x - 1$
H. $y = \frac{1}{3}x - 1$
J. $y = \frac{1}{3}x + 1$
K. $y = \frac{1}{2}x - 1$

Items 6 and 20 since they measure in the same direction, but it would not be reasonable to compare the difficulty of Items 20 and 34 because they are measuring in quite different directions.

Items that measure best in a particular direction can be combined together to form a test that will operate as if it were measuring a single dimension. For example, if Items 4, 13, 16, 25, and 33 were selected as a single subtest, that subtest would operate as a unidimensional subtest because all of the items measure the same composite of θ_1 and θ_2. All of these items can also be ordered in difficulty because they all measure the same composite of abilities.

The concept of MID can also be used to select items to measure a person at a particular $\boldsymbol{\theta}$ point. The MID definition emphasizes that, when selecting an item, the direction must be considered as well as the point of maximum discrimination. If information is desired for all θ dimensions, the item must be selected so that the slope of the IRS is nonzero for any direction parallel to a θ axis. This means that the item direction must not be parallel to any axis. Item 36 is such an item. If one dimension is of interest to the exclusion of the others, items should be selected that have directions parallel to the dimension of interest.

Conclusions

The definition of item difficulty proposed here describes the difficulty of an item as a direction and a distance in the complete latent space. The use of the definition was demonstrated using the multivariate extension of the two-parameter logistic model, but the definition is sufficiently general that it can be used with any model that yields probabilities that increase monotonically with an increase in ability on any dimension.

References

McKinley, R. L., & Reckase, M. D. (1983a). *An extension of the two-parameter logistic model to the multidimensional latent space* (Research Report ONR 83-2). Iowa City IA: The American College Testing Program.

McKinley, R. L., & Reckase, M. D. (1983b). MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. *Behavior Research Methods and Instrumentation, 15*(3), 389–390.

Reckase, M. D., & McKinley, R. L. (1983, April). *The definition of difficulty and discrimination for multidimensional item response theory models*. Paper presented at the meeting of the American Educational Research Association, Montreal, Canada.

Traub, R. E., & Lam, Y. R. (1985). Latent structure and item sampling models for testing. *Annual Review of Psychology, 36,* 19–48.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST user's guide*. Princeton NJ: Educational Testing Service.

Acknowledgments

This research was supported by Contract Number N00014-81-K0817 from the Personnel and Training Research Programs of the Office of Naval Research.

Author's Address

Send requests for reprints or further information to Mark D. Reckase, ACT, P.O. Box 168, Iowa City IA 52243, U.S.A.