The Use of Unidimensional Parameter Estimates of Multidimensional Items in Adaptive Testing

Terry A. Ackerman
University of Illinois

This study investigated the effect of using multidimensional items in a computerized adaptive test (CAT) setting which assumes that all items are unidimensional. Previous research has suggested that the composite of multidimensional abilities being estimated by a unidimensional IRT model is not constant throughout the entire unidimensional ability scale (Reckase, Carlson, Ackerman, & Spray, 1986). Results of this study suggest that unidimensional calibration of multidimensional data tends to "filter out" the multidimensionality. Items that measured a θ_1, θ_2 composite similar to the composite of the calibrated unidimensional θ scale had larger estimated unidimensional discrimination values. These items thus had a greater probability of being administered in a CAT where only the most informative items are selected. Results also suggest that if a CAT item pool contains items from several content areas measuring dissimilar θ_1, θ_2 composites, different unidimensional abilities may receive disparate proportions of items from the various content areas. *Index terms: adaptive testing, item response theory, multidimensionality, parallel tests, test construction.*

Most item response theory (IRT) models assume that an examinee's test performance can be explained by a single ability or latent trait—that is, an examinee's position in the latent ability (θ) space can be determined by measuring a single θ dimension. However, this assumption may be rarely met because of the many cognitive factors that may account for an individual's response to an item (Traub, 1983). For a group of individuals, it is doubtful that a single cognitive skill or constant combination of skills would be used by each person to respond to a single item. Thus, it is even more doubtful that the assumption of unidimensionality would be met for a group of individuals responding to an entire test.

Reckase, Carlson, Ackerman, and Spray (1986) showed that the composite of θs did not remain consistent throughout the estimated unidimensional θ scale for the measurement of generated two-dimensional data in which difficulty and dimensionality were confounded (i.e., easy items measured primarily the first dimension and difficult items measured primarily the second dimension). Specifically, they reported that for the particular confounding of θ and difficulty used, the examinees in upper LOGIST estimated θ deciles differed mainly on θ_2, whereas those in the lower $\hat{\theta}$ deciles differed mainly on θ_1.

If these results are generalizable to real achievement test items, the application of computerized adaptive testing (CAT) could be profoundly affected. When an adaptive test item pool is composed of items that require different composites of ability to answer correctly, low- and high-ability individuals may be administered two sets of items that measure completely different combinations of skills.

Weiss and Suhadolnik (1982) investigated the robustness of adaptive testing to the violation of the unidimensionality assumption. Using a factor-analytic approach, they generated item parameters that mirrored the multidimensional factor structure of the Armed Services Vocational Aptitude Battery's General Science test. They simulated adaptive tests using 45 different item pools, each a variation of the General Science

test's four-factor structure. Item selection was based on maximum information at the current estimated θ. Item information was determined using item parameter estimates provided by the items' first factor loadings. For each selected item, the probability of a correct response was computed separately for each of the four dimensions. These were then combined by weighting the probabilities by their squared factor loadings. Finally, the combined probability was compared to a random deviate from a uniform 0,1 distribution to determine a correct response. The study focused on the relationship between the estimated θ and the true θ generated for the first dimension.

The results of Weiss and Suhadolnik (1982) suggest that as multidimensionality increases, the estimated θs deviate more from their true values. However, they also concluded that increasing test length can overcome the "effects of multidimensionality with one or two factors beyond the first that account for up to one-fourth the variance of the first factor" (p. 278).

The present study differs from Weiss and Suhadolnik (1982) in several ways. First, multidimensionality was induced here into the adaptive testing process with the compensatory IRT model developed by Reckase (1985), rather than using a factor-analytic approach. Second, the effect of multidimensionality of item pools on the mapping of the ability space onto the unidimensional θ scale was examined. This differed from Weiss and Suhadolnik (1982), who examined how differing degrees of multidimensionality affected the relationship between $\hat{\theta}$ and the true first factor θ. Third, the issue of different θ levels receiving percentages of items from different content categories was investigated.

When two-dimensional abilities are mapped onto a unidimensional scale, the mapping is not one-to-one: several θ_1,θ_2s are mapped into the same unidimensional $\hat{\theta}$. From a visual perspective, the mapping of θ_1,θ_2 onto $\hat{\theta}$ produces a "plane" of estimated unidimensional θs, which often intersects the two-dimensional ability plane. The estimated unidimensional ability plane is defined by the multiple regression of the estimated

θs onto the θ_1,θ_2 plane. The orientation of this plane can best be examined by plotting the equi-unidimensional θ contour projections onto the θ_1,θ_2 plane. By studying the alignment of these contours with the θ_1,θ_2 axes, it can be determined whether one ability is being measured more accurately than the other.

Figure 1 shows an example that illustrates the spatial relationship between a two-dimensional θ space and the $\hat{\theta}$ regression plane. The regression equation for this example is $\hat{\theta} = .50\theta_1 + .50\theta_2$. The positive portion of the $\hat{\theta}$ plane lies above the θ plane, and the negative portion lies below. Although it is characterized as a plane of $\hat{\theta}$s, all locations on this plane that are the same perpendicular distance from the θ plane have the same $\hat{\theta}$. These distances or "heights" are revealed when the contour plot of the $\hat{\theta}$ plane is projected onto the θ plane (Figure 2). The alignment and spacing of the contours in this example verify that both dimensions are being measured with the same degree of precision.

Figure 1
The Regressed $\hat{\theta}$ Plane in Relation to the True θ Plane
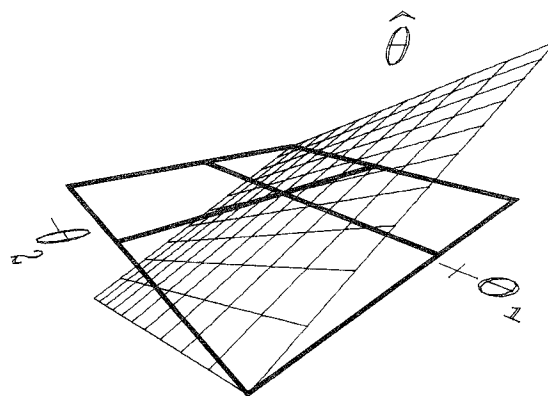
The unidimensional item parameter estimates of multidimensional items are thought to be a function of the orientation of the estimated unidimensional θ plane in the multidimensional θ space. This alignment or orientation has been investigated by Wang (1986). She demonstrated analytically how the alignment of the plane of unidimensional estimates is a function of both

Figure 2
Projected Contours of the $\hat{\theta}$ Plane Onto the θ Plane
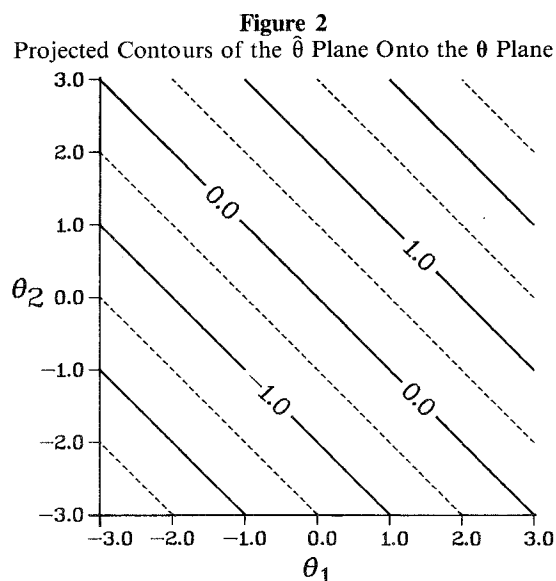
the underlying multidimensional ability distribution and the multidimensional characteristics of the items. The orientation of the plane of unidimensional estimates was characterized by what Wang called the "reference composite." The reference composite is equivalent to the regression equation of θ_1, θ_2 onto $\hat{\theta}$. The θ contours will always be orthogonal to the line having an angle equal to $\tan^{-1}(\beta_2/\beta_1)$ with the θ_1 axis (where β_1 and β_2 are the standardized regression coefficients). This angle can be termed the reference angle.

One conclusion that follows from Wang's work is that the orientation of the reference composite is influenced by the pattern of multidimensional information over the θ_1, θ_2 plane (i.e., if a test provides more information about θ_1 in the two-dimensional case, a unidimensional calibration of response data from this test would tend to "orient" the unidimensional $\hat{\theta}$ plane along the θ_1 axis). Contour lines of the $\hat{\theta}$ plane, when projected on the θ plane, will run obliquely to the θ_1 axis and somewhat parallel to the θ_2 axis. If a dataset provides uniform information throughout the θ plane, the measurement precision of both dimensions will be equal and the projected contours of the unidimensional plane

will have a reference angle of a 45° line (see Figures 1 and 2). However, if the individual items differ in the composite of abilities needed for a correct response, tests composed of different items could have different unidimensional orientations in relationship to the θ plane. How the calibrated unidimensional scale is positioned in the plane may affect the way different locations in the plane are mapped onto the unidimensional θ scale.

Samejima (1977) suggested that unidimensional tests are weakly parallel if they provide the same amount of information at each point throughout the θ scale. The logical multidimensional extension would be that two tests are weakly parallel if the composite of abilities being measured by each test is being assessed with the same degree of precision for all points throughout the ability space.

Purpose

The underlying issue investigated here was whether persons at different locations along the estimated unidimensional θ scale in a CAT administration of two-dimensional items would receive parallel tests (i.e., whether tests would vary in the θ_1, θ_2 composite of ability being measured and/or the degree of measurement precision would not be equal across tests). Two research questions were examined. The first concerned the orientation of the contours of the plane of unidimensional $\hat{\theta}$s projected onto the θ plane, to determine whether the orientation would differ as the θ_1, θ_2 composite of items in a test changed. The second research question, based on the findings of previous research by Reckase et al. (1986), concerned whether individuals in a CAT administration at different ability levels on the mapped unidimensional scale would receive disparate proportions of item content requiring different θ_1, θ_2 combinations for a correct response.

Two experiments were conducted. The first experiment examined only the first research question using generated data. The second experiment examined both questions using real data in which difficulty was suspected to be confounded with

dimensionality. Two different test formats were used in each experiment. The first was an adaptive test format (CAT), and the second was an administration of the complete item pool (CPA). For both formats, two simulations were conducted to establish the projected unidimensional scale orientation onto the θ plane scale. To investigate the second question, a simulation was conducted in the second experiment using the CAT format to determine if different ability levels along the mapped unidimensional scale received tests composed of dissimilar proportions of six different content categories, each providing different information about the two-dimensional θ plane.

Experiment 1

Method

Generation of item response data. Using the two-dimensional compensatory IRT model, an item pool of 100 items was created. For the purposes of this study, the multidimensionality issue was not confounded with the guessing issue; thus the two-parameter multidimensional logistic (M2PL) compensatory model was selected (Reckase, 1985). In this model, the probability of a correct response to item i by person j is given as:

$$P(X_{ij} = 1|\mathbf{a}_i, d, \theta_j) = 1 + \exp{-(a_1\theta_1 + a_2\theta_2 + d)}, \quad (1)$$

where X_{ij} is the response to item i by person j,
$\mathbf{a}_i$ is a vector of item discrimination parameters,
d_i is the multidimensional difficulty of item i, and
θ_j is a vector of person parameters.

Item parameters for 100 items were generated by selecting multidimensional item discrimination values (Reckase, 1985) randomly from a β distribution ($\alpha = .11$, $\beta = .11$). The selection of discrimination values from this bimodal distribution and the pairing of them with multidimensional difficulty parameters [randomly selected from a uniform U(0,1) distribution] created uniform information in each direction throughout the two-dimensional θ plane in the region $3 < \theta_1 < 3$, $-3 < \theta_2 < 3$. The length of each item vector (see Figure 3) represents the amount

Figure 3
Vectors for the Generated Item Parameters Representing the Direction and Distance to the Point of Maximum Discrimination
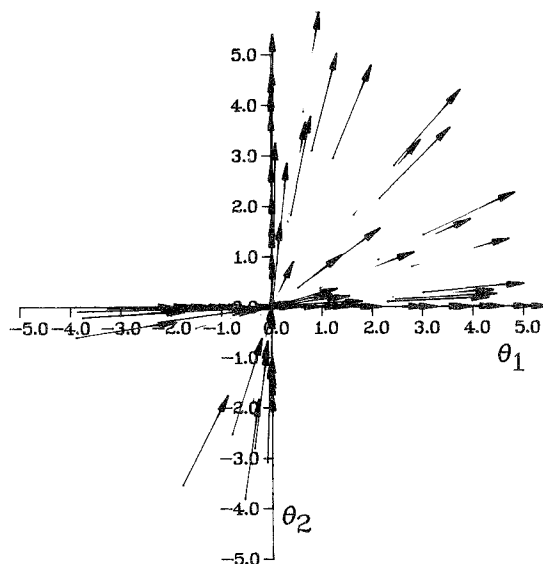
of discrimination for a given item. The $p = .5$ equiprobability contour for each item is orthogonal to the base of the item vector. To achieve uniform information throughout the θ plane, the majority of items had to measure predominantly either θ_1 or θ_2. The summative effect of combining information of items that measured mostly θ_1 or θ_2 also provided equal amounts of information in a direction of 45° from the θ_1 axis.

In the first simulation, 2,000 θ_1, θ_2s were randomly selected from a bivariate normal distribution, $N(0, \Sigma)$ where $\Sigma = I$. Using the M2PL model, response vectors were generated for each simulee. The response vectors were subsequently calibrated to fit a unidimensional two-parameter logistic (2PL) IRT model using the IRT calibration program LOGIST (Wingersky, Barton, & Lord, 1982). The two test formats (CAT and CPA) were simulated for the 2,000 θ_1, θ_2s previously selected using both the multidimensional item parameters and their corresponding unidimensional estimates.

Test administration. In the CAT simulation, the initial unidimensional estimate of θ was 0.

Items were selected using their unidimensional information values. Those items that provided the most information at the current $\hat{\theta}$ were selected. Then the probability of a correct response was computed using the known two-dimensional item parameters and the preselected θ_1, θ_2 according to the M2PL model. This probability was then compared to a randomly generated threshold value from a U(0,1) distribution to yield a correct or incorrect response, and the unidimensional $\hat{\theta}$ was updated. This iterative CAT process was continued until either 20 items had been administered, or the sum of the unidimensional item information reached 10 at the current $\hat{\theta}$.

The CPA simulation was identical to the CAT test, except that no limit was placed on the amount of unidimensional information, and the maximum test length was increased to 100. By relaxing the stopping rule criteria in such a manner, each simulee was administered the entire item pool.

Measurement accuracy.   For each type of administration, a multiple regression was conducted in which the 2,000 $\hat{\theta}$s were regressed onto the generated θ_1, θ_2s. Contour plots of the $\hat{\theta}$ plane were then constructed for each simulated administration format. The orientation of the contours in relationship to the θ_1, θ_2 axes provided information about the measurement accuracy of each dimension, which was then compared across each testing format.

The mapping of the two-dimensional θ space onto the unidimensional scale can best be understood by examining the response surface of three selected compensatory items and the corresponding probabilities of correct response for three distinct points in the θ plane. In Figure 4, the contour plots show the lines of equiprobability for a correct response for the three equally discriminating items. Three selected θs, $A(2,0)$, $B(0,0)$, and $C(0,2)$, are indicated on each contour.

Item 1 (Figure 4a) discriminates or provides information only along θ_2. If a test were composed of items identical to Item 1, multidimensional abilities B and C would receive the same
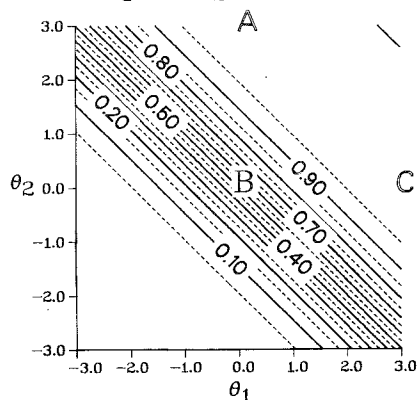
Figure 4
Contour Plots of Three Compensatory Items
(a) An Item That Differentiates Along θ_2 Only
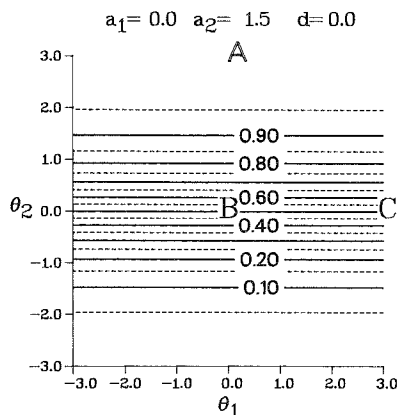
(b) An Item That Differentiates Along θ_1 Only
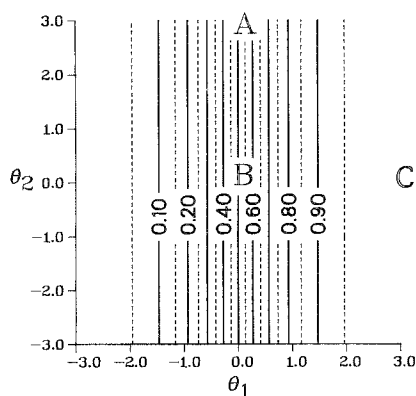
(c) An Item That Differentiates Equally on θ_1 and θ_2

unidimensional θ estimate; however, this estimate would be less than the estimated θ for ability A. The reference angle in this case is 90°. Item 2 (Figure 4b) provides information only along θ_1. In a test composed of items of this type, multidimensional abilities A and B would receive the same unidimensional θ estimate, although point C would be assigned to have a higher ability. The reference angle in this case is 0°. Item 3 (Figure 4c) represents an item that discriminates equally well on both θ_1 and θ_2. If a test were to be composed of items of this type, A and C would be estimated to have the same unidimensional θ but B would be assigned a much lower $\hat{\theta}$. The compensatory nature of the test becomes evident in this case, when the reference angle is at 45°—that is, examinees who have average ability on one dimension can compensate by having a high ability on the other dimension. Thus, points A and C are mapped into the same $\hat{\theta}$.

By examining the relationship of the projected contours of the plane of estimated unidimensional θs and the corresponding reference angles, it can be determined whether a test is measuring mostly θ_1, mostly θ_2, or both θ_1 and θ_2 equally well. The same analysis can be conducted with other triads of points throughout the two-dimensional ability plane.

Results

The following standardized functions were obtained for each administration format:

$$\hat{\theta}_{\text{CPA}} = .78\theta_1 + .58\theta_2 \tag{2}$$

and

$$\hat{\theta}_{\text{CAT}} = .88\theta_1 + .30\theta_2 \ . \tag{3}$$

Contour plots, illustrating the orientation of each format's estimated θ plane, are shown in Figure 5, which displays the equi-unidimensional estimated θ contours projected onto the θ_1,θ_2 plane. Several interesting comparisons can be made by examining these plots. First, the CPA reference angle is 36.6°, compared to 18.8° for the simulated CAT. Therefore, the CAT process discriminated better between θ_1 abilities and θ_2

Figure 5
Contour Plots of the Estimated $\hat{\theta}$ Plane Projected
on the θ Plane for Simulations Using
Generated Item Parameters

(a) CPA Simulations

(b) CAT Simulations
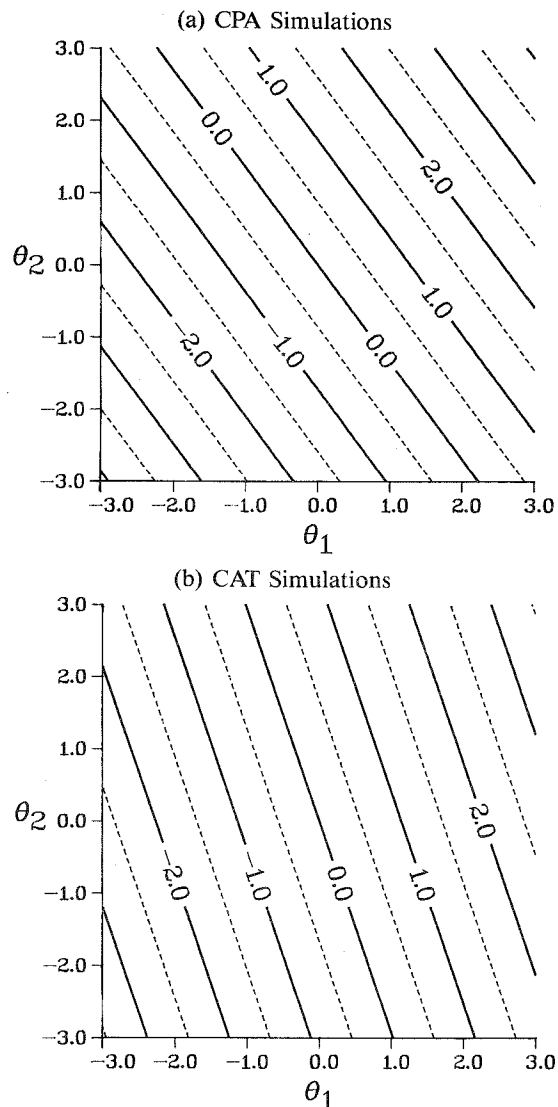
abilities than did the CPA administration. This implies that items that discriminate better on θ_1 had larger calibrated univariate discrimination parameter estimates and were selected more often in the simulated CAT. Thus, it can be assumed that the LOGIST calibration process tended to emphasize the first dimension more than the second

dimension, and estimated larger discrimination values for items measuring predominantly θ_1.

The estimated multiple regression equations for both administrative formats used only first-order terms. Higher-order terms (i.e., θ_1^2, θ_2^2, or $\theta_1\theta_2$) did not explain a significant amount of the $\hat{\theta}$ variance.

The projected CPA contours (Figure 5a) are slightly closer together, which suggests that the CPA test did a better job of discriminating between individuals than did the CAT format. This result was to be expected, because of the larger discrimination power obtained by administering all the items. Each CAT simulated test could involve, at most, only one-fifth as many items as the CPA test. It was therefore expected that the CAT could not distinguish between the multidimensional abilities as well as the CPA format. However, because the CPA contour lines are only slightly closer than the CAT contours (Figure 5b), it can be concluded that the items not selected in the CAT simulation did not provide as much information within the range of the simulated abilities as might be expected.

Discussion

These results suggest several areas of concern for the practitioner. Even though a group of items considered collectively provides uniform information throughout a two-dimensional ability plane, this does not guarantee that tests containing smaller subsets will also provide uniform information. The orientation of the projected unidimensional scale appears to be a function of the informational structure of the items that comprise the test.

The LOGIST calibration process tended to emphasize only one of the dimensions, despite the uniformity of multidimensional information throughout the θ plane. The more similar an item's θ_1,θ_2 composite was to the composite of the LOGIST orientation, the larger the estimated unidimensional discrimination value was for the item. Items that predominantly measure abilities other than those emphasized by the LOGIST orientation will have very small unidimensional

discrimination parameter estimates. This result is critical when viewed from a CAT perspective.

Because the calibration orientation ultimately determines the size of the estimated discrimination parameters, it will subsequently determine which items will have a high probability of being selected in a unidimensional CAT administration. Thus, the unidimensional calibration process tends to mask the multidimensional nature of the pool by emphasizing only a single θ_1,θ_2 composite. However, more research is needed to investigate whether the θ_1,θ_2 composite is being measured consistently throughout the entire unidimensional ability scale.

One weakness to consider when comparing the difference between the two simulated administration formats is that the standard errors of the CPA unidimensional θ estimates were considerably smaller than those estimates obtained in the CAT administration. This was simply an artifact of the number of items administered in each format.

Experiment 2

Method

In this study, the simulated data were based on the characteristics of actual test data. The item pool was created from the ACT Assessment Math Usage Test Form 26A. The test contained 40 multiple-choice items covering six content areas: Geometry (G, 8 items), Number and Numeration Systems (NNS, 4 items), Intermediate Algebra (IA, 8 items), Algebraic and Arithmetic Operations (AAO, 4 items), Arithmetic and Algebraic Reasoning (AAR, 14 items), and Advanced Topics (AT, 2 items) (American College Testing Program, 1986).

Item response generation. Using 3,000 randomly selected response vectors from a national administration, two-dimensional item parameters were estimated using the multidimensional IRT estimation program MIRTE (Carlson, 1987). The calibrated compensatory parameters were then used to expand the 40-item set so that the six content areas each contained 16 items. For example,

the eight geometry items in the original 40-item test were simply replicated to produce a total of 16 items. The purpose of expanding the item pool was twofold: first, to increase the size of the CAT pool; and second, to assure that each content area had the same number of items in the pool.

The same 2,000 θ_1, θ_2 pairs that were used in Experiment 1 were used to create response vectors for the 96 items using the M2PL compensatory model. The simulated response data were then calibrated using LOGIST to obtain the two-parameter IRT item parameter estimates. Again two different simulations were conducted, one for each administration format. The CPA and the CAT simulations were conducted as described in Experiment 1.

Item-content effects.   In addition to creating the contour plots, an additional simulation was conducted at each of seven equally spaced θ values from -3 to +3 on the unidimensional scale. At each of these unidimensional θs, a CAT test was simulated 100 times, using two-dimensional item and θ parameter estimates. Seven two-dimensional θ_1, θ_2 pairs, which mapped onto the seven unidimensional $\hat{\theta}$s, were arbitrarily selected with the multiple regression equation computed for the CAT administration format. During each simulation, the content of the administered items was recorded. The proportions of each type of content were computed and compared at each selected $\hat{\theta}$.

Results

Multidimensional information was examined for both the entire pool and the individual content categories. The item pool did not provide uniform information throughout the two-dimensional θ plane. The G, NNS, IA, and the AAO items all provided more information along the second θ dimension, albeit at different places in the θ plane. The AAR items appeared to be more discriminating along the first θ dimension. The information vector plots for the IA and AAR items are shown in Figure 6. In these figures, information vectors are displayed for 49 selected points in the θ plane. The vectors are plotted in

Figure 6
Multidimensional Test Information Vectors for 49 Selected Points in the θ Plane

(a) IA Items

(b) AAO Items

(c) AT Items
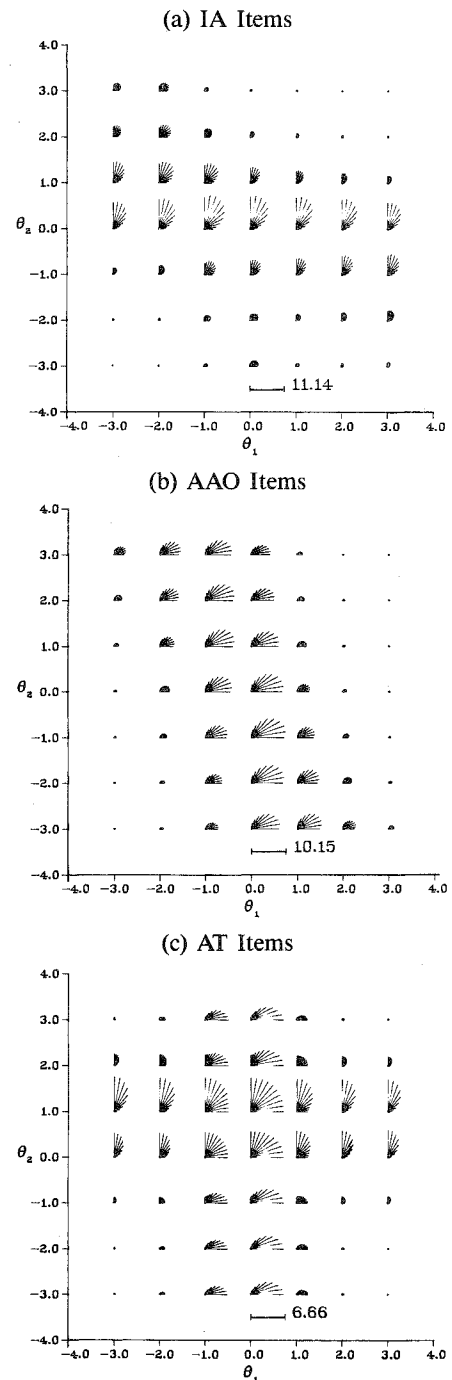
10° increments from 0° to 90°. The length of each vector represents the height of the multidimensional test information surface at θ_1, θ_2 in the direction of the selected angle (Reckase, 1986).

After examining the content of these item types, it became clear that the first dimension could be characterized as a verbal reasoning ability, because most of the AAR items were "story" or word problems. In responding to these types of questions, the examinee must first translate the written text into an algebraic framework before computing a solution. The IA, G, AAO, and NNS items generally involved some form of algebraic computation. Thus, the second dimension was defined as an algebraic symbol manipulation ability.

Interestingly, the hypothesized characterization of the two dimensions was confirmed by the AT items. Although there were only two AT items in the original 40 items, one of these was a story problem, and the other involved algebraic manipulation. The multidimensional information plot for these items (Figure 6c) revealed that the two items were providing the most information in a band along each axis. It was assumed that the horizontal band reflected the information provided by the story problem, and that the vertical band reflected the information provided by the algebraic problem.

For each content area, the unidimensional test information functions were computed and compared. The pool information function as well as the individual content information functions are displayed in Figure 7. The AAO content area provided the highest level of information, whereas the NNS items provided the least. Compared to their multidimensional counterparts, the univariate content information functions provide a striking contrast, revealing a high degree of similarity.

The plots of the projected contours for the CAT and CPA administrations are shown in Figure 8. The multiple regression functions, obtained by regressing $\hat{\theta}$ onto the known θ_1, θ_2 for each format, were

Figure 7
The Total Pool and Individual Content Area
Unidimensional Information Functions
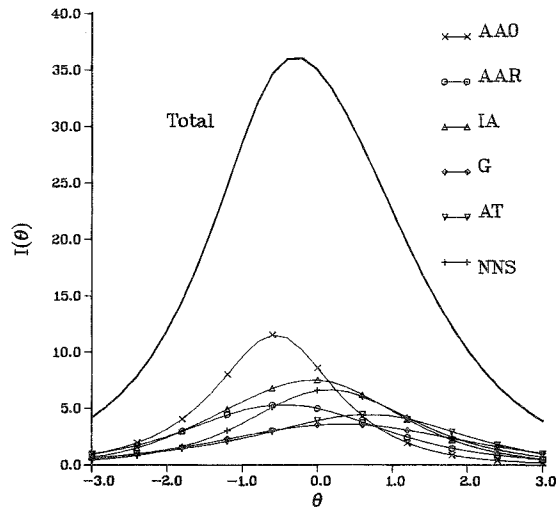
$$\hat{\theta}_{CPA} = .66\theta_1 + .71\theta_2 \tag{4}$$

and

$$\hat{\theta}_{CAT} = .49\theta_1 + .82\theta_2 \; . \tag{5}$$

The orientation of the contours of the estimated unidimensional scale for each format was similar to the results obtained in Experiment 1. The CAT reference angle (Figure 8b) was 59.1°, which suggested that the items administered in this format were discriminating better along θ_2 than along θ_1. Both dimensions appeared to be assessed equally well in the CPA format (Figure 8a), which had a reference angle of 47.1°. Despite more items administered in the CPA format, the contour lines for both plots appear to be spaced similarly (i.e., even though more items were administered in the CPA format, the overall degree of discrimination was not noticeably improved).

The proportions of each content administered at the seven estimated unidimensional abilities are displayed in Table 1. A clear shift in the item content is evident. No NNS items were selected at any of the seven θ values. This artifact may be partly due to the LOGIST orientation in the θ plane (i.e., if the NNS items were measuring mostly θ_2 and

Figure 8
Contour Plots of the Estimated $\hat{\theta}$ Plane Projected on
the θ Plane for Simulations Using Estimated
Item Parameters

(a) CPA Simulations

(b) CAT Simulations
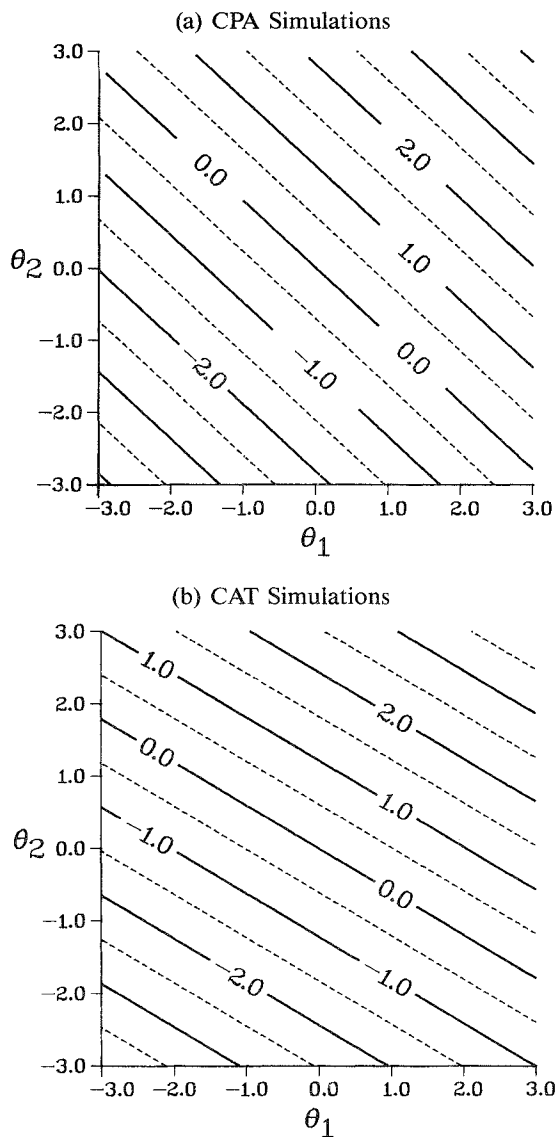
the LOGIST orientation tended to emphasize
measurement along θ_1, the unidimensional dis-
crimination parameters would be estimated to be
very low). Consequently, no NNS was selected in
the CAT administration. This is also supported
by the fact that the unidimensional NNS con-

Table 1
The Proportion of Each Content Area
Sampled in the Multidimensional CAT
Simulation at Selected θ Levels

Content Area	θ Level						
	−3	−2	−1	0	1	2	3
AAR	.16	.15	.13	.14	.07	.03	.01
NNS	.00	.00	.00	.00	.00	.00	.00
IA	.28	.26	.23	.27	.33	.36	.38
G	.06	.12	.19	.26	.27	.24	.19
AAO	.50	.57	.45	.32	.23	.18	.16
AT	.00	.00	.00	.01	.10	.19	.26

tent information curve (Figure 7) was the lowest
of the six content areas.

Only the IA items appear to be represented by
a constant percentage across the θ scale. Fifty per-
cent of the items administered at $\hat{\theta} = -3$ were
AAO items, whereas only 16% of the items ad-
ministered at $\hat{\theta} = 3$ were AAO items. No AT items
were administered below a $\hat{\theta}$ value of 0, yet over
one-fourth of the items administered at $\hat{\theta} = 3$
were AT items.

Conclusions

The results of this study strongly suggest that
the concerns over persons of different ability lev-
els receiving different content in adaptive testing
are valid. In Experiment 2, the percentages of the
six content areas administered throughout the
univariate scale differed noticeably as a function
of θ. It would be interesting to see how dramatic
the shift in $\hat{\theta}$ would be if the CAT was ad-
ministered in a way that would control the num-
ber of items administered from each content area.
Because each CAT test may differ in length, this
problem is not easily solved.

The orientation of the unidimensional scale
in relationship to the two-dimensional θ plane ap-
pears to be a function of the multidimensional
composition of the items administered in the test.
This is the same conclusion reached by Wang
(1986). The item selection process in a CAT ad-
ministration tends to make the unidimensional
scale orientation similar to that obtained in the
original unidimensional calibration process. This
filtering of the dimensionality occurs because

unidimensional discrimination estimates increase as an item's θ_1, θ_2 composite gets closer to the composite of the unidimensional calibrated scale. However, the effectiveness of the items for measuring multiple dimensions is misrepresented.

Multidimensional information appears to be quite helpful in revealing suspected measurement differences between different content areas. Although unidimensional information functions may display a high degree of similarity, their multidimensional counterparts can be quite dissimilar. Multidimensional differences can be used to identify and distinguish between the types of skills required to solve various types of problems. Such information could become an important step in the test development process.

This study illustrated graphically—with contour plots—how different sets of multidimensional items can affect the mapping of the two-dimensional θ plane onto the unidimensional scale. For both the generated data and the quasi-real data, there appeared to be differences in the degree to which each ability dimension was being measured for different θ_1, θ_2 points in the ability plane. It might be argued that this could only occur with a strictly unidimensional item pool.

Results of both experiments tend to support one of the findings of Weiss and Suhadolnik (1982), who suggested that the effects of multidimensionality can be overcome (i.e., canceled out) by increasing the number of items. In both experiments, the CPA format resulted in a reference angle that was closer to 45°. This would imply that as the length of the test increases, both dimensions are measured approximately equally.

The present results do not support the confounding effect noted by Reckase et al. (1986). Even though difficulty was confounded with dimensionality in the last experiment, no higher-order terms (e.g., θ_1^2, θ_2^2) explained enough of the $\hat{\theta}$ variance to enter the equation. Thus, $\hat{\theta}$ appears to be best explained by a linear combination of θ_1 and θ_2 through the two-dimensional ability plane. These results also suggest that more research is necessary in order to understand the purifying process that occurs when multidimensional data are calibrated to fit unidimensional IRT models. Practitioners need to determine if such filtering is really beneficial, or if it is more advantageous from an instructional point of view to work from the multidimensional perspective.

References

American College Testing Program. (1986). *The ACT Assessment test preparation reference manual for teachers and counselors.* Iowa City IA: Author.

Carlson, J. E. (1987). *Multidimensional item response theory estimation: A computer program.* Unpublished manuscript.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9,* 401–412.

Reckase, M. D. (April, 1986). *The discriminating power of items that measure more than one dimension.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco CA, U.S.A.

Reckase, M. D., Carlson, J. E., Ackerman, T. A., & Spray, J. A. (June, 1986). *The interpretation of unidimensional IRT parameters estimated from multidimensional data.* A paper presented at the annual meeting of the Psychometric Society, Toronto, Ontario, Canada.

Samejima, F. (1977). A use of the information function in tailored testing. *Applied Psychological Measurement, 1,* 233–247.

Traub, R. E. (1983). A priori considerations in choosing an item response model. In R. K. Hambleton (Ed.), *Applications of item response theory* (pp. 57–70). Vancouver, British Columbia: Educational Research Institute of British Columbia.

Wang, M. M. (1986). *Fitting a unidimensional model to multidimensional item response data.* Paper presented at the annual meeting of the Office of Naval Research Contractor's Meeting, Knoxville TN, U.S.A.

Weiss, D. J., & Suhadolnik, D. (1982). Robustness of adaptive testing to multidimensionality. In D. J. Weiss (Ed.), *Proceedings of the 1982 item response theory and computerized adaptive testing conference.* Minneapolis MN: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory.

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

Author's Address

Send requests for reprints or further information to
Terry A. Ackerman, Educational Psychology Depart-
ment, University of Illinois, Champaign IL,
61820-6990, U.S.A.