ACT Research Report Series 87-13

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

Terry Ackerman

September 1987

ACT

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243

THE USE OF UNIDIMENSIONAL ITEM PARAMETER ESTIMATES OF MULTIDIMENSIONAL ITEMS IN ADAPTIVE TESTING

Terry A. Ackerman

ABSTRACT

The purpose of this study was to investigate the effect of using multidimensional items in a computer adaptive test (CAT) setting which assumes a unidimensional IRT framework. 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 univariate calibration of multidimensional data tends to "filter" out the multidimensionality. The closer an item's multidimensional composite aligns itself with the calibrated univariate ability scale's orientation, the larger the estimated discrimination parameter. If CAT item selection is based upon the amount of information an item provides, items requiring similar (θ_1, θ_2) composites will most often be selected.

These results further imply that in a CAT different abilities throughout the θ_1, θ_2 plane could receive sets of items that discriminate between θ_1 and θ_2 to different degrees. Also, different abilities along the mapped univariate scale, could receive tests having different proportions of item content.

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

Most item response theory 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 space can be determined by measuring a single ability dimension. However, one might suspect that this assumption is rarely met because there are 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. It is even more highly suspect that this assumption of unidimensionality would be met for a group of individuals responding to an entire test.

Reckase, Carlson, Ackerman, & Spray (1986) have shown that for generated two dimensional data, where difficulty and dimensionality of the items are confounded (e.g., easy items measure only ability 1 and difficult items measure only ability 2), the unidimensional ability estimation scale is related to different composites of the two abilities at different points on the unidimensional ability scale. Specifically, they reported that for the particular confounding of ability and difficulty used, the examinees in upper LOGIST estimated ability deciles differed mainly on θ_2 while those in the lower deciles differed mainly on θ_1.

If these results are generalizable to real achievement test items, it could have a profound effect on the application of computer adaptive testing (CAT). If an adaptive test item pool is composed of items which require different composites of ability to answer correctly, low ability and high ability individuals may be administered two sets of items that measure completely different combinations of skill.

The unidimensional item characteristics are thought to be a function of the alignment of the ability scale in the multidimensional ability space. This alignment or orientation is strongly influenced by the pattern of the multidimensional test information over the θ_1, θ_2 plane. That is, if more information (e.g. more discrimination) is provided along the θ_1 axis, calibration using a univariate IRT model would orient the univariate ability scale along the θ_1 axis. If a dataset provided uniform information throughout the ability plane, both dimensions should be represented equally well and the univariate scale should be mapped at a 45° angle between the θ_1, θ_2 axes. However, if the individual items differed in the composite of abilities needed for a correct response, tests composed of different items could have different orientations in the θ_1, θ_2 plane. How the calibrated univariate scale is positioned in the plane may affect how different locations in the plane are mapped onto the scale.

Samejima (1978) has suggested that univariate tests are truly parallel if they provide the same amount of information on each point throughout the ability scale. The logical multidimensional extension would be that two tests are parallel if the composite of abilities required for a correct response is the same for all points in the ability space. The primary focus of this study was to examine the concerns of Reckase and colleagues that persons at different locations along the theta scale would not receive parallel tests or items having the same ability composites.

Three hypotheses were examined in this study. First, it was believed that the orientation of the univariate theta scale in the two dimensional ability plane would differ as the composite of items in the test changed. A second part of this study was to briefly examine how the θ_1, θ_2 plane was mapped onto the univariate scale. It was suspected that differences in test administration format could change the way in which θ_1, θ_2 combinations are mapped onto the univariate scale. Third, based upon the findings of previous research by Reckase, et al. (1986) it was hypothesized that individuals with different ability levels on the mapped univariate ability scale would receive items from a CAT that would require different θ_1, θ_2 combinations for a correct response.

To verify these hypotheses, two experiments were conducted. The first experiment examined the hypotheses with generated data, while the second used real data in which difficulty was known to be confounded with dimensionality. To confirm the first two hypotheses two different test formats were used in each experiment. The first was an adaptive test format (CAT); the second was an administration of the entire item pool (CPA). For both CPA formats, two simulations were conducted, one to establish the univariate scale orientation in the θ_1, θ_2 plane, and one to study the mapping of the θ_1, θ_2 plane onto the univariate scale. To verify the third hypothesis a third simulation for the CAT format only was conducted to determine if abilities along the mapped univariate scale received tests composed of different items.

Experiment 1

Method

Generation of Item Response Data

Using a two dimensional IRT model, a test item pool of 100 items was created with multidimensional item discrimination, MDISC values (See Reckase, 1986) randomly selected from a beta distribution (α = .11, β = .11). The multidimensional difficulty parameters were randomly selected from a uniform

U(0, 1) distribution. By selecting item parameters in this manner, multidimensional information (Reckase, 1986) remained relatively constant throughout the ability plane. The multidimensional information is shown in the plot of the test information given in Figure 1. Vectors at 10 degree increments for 49 selected points are presented. The length of each vector represents the height of the multidimensional test information surface at the (θ_1, θ_2) point in the direction of the selected angle. The length of the vectors are about equal at each angle at each point, indicating uniform information.

A plot of the item vectors is shown in Figure 2. Each item vector represents the distance and direction from the origin to the point of maximum discrimination for a given item. To achieve uniform information, the majority of items had to measure predominantly either θ_1 or θ_2. This is shown in Figure 2. The majority of item vectors are either positioned close to the θ_1 or close to the θ_2 axis.

In the first simulation, 2,000 (θ_1, θ_2) combinations were randomly selected from a bivariate normal distribution, N $(\underset{\sim}{0}, \sum)$ where $\sum$ was the identity matrix, to simulate response vectors to the set of 100 selected item parameters using the compensatory multidimensional item response theory model (see Reckase, 1985). In this model the probability of a correct response to item i by person j is given as:

$$P(\underline{X}_{ij} = 1 | \underline{a}_i, d, \underline{\theta}_i) = \frac{e^{\underline{a}_i \underline{\theta}_j + d_i}}{1 + e^{\underline{a}_i \underline{\theta}_j + d_i}} \tag{1}$$

where $\underline{X}_{ij}$ is the response to item i by person j,

$\underline{a}_i$ is a vector of item discrimination parameters,

d_i is an item parameter related to the difficulty of an item, and

$\underline{\theta}_j$ is a vector of person parameters.

Procedure

The generated response vectors were calibrated to a unidimensional two-parameter IRT model using LOGIST (Wingersky, Barton, & Lord, 1982). Using these item parameter estimates, the two test formats (CAT and CPA) were simulated for 1000 (θ_1, θ_2) ability combinations from the $N(\underset{\sim}{0}, \Sigma)$ distribution mentioned previously. The purpose of selecting these two testing formats was to determine how each format mapped the univariate scale onto the two dimensional ability plane.

In the CAT simulation, the initial unidimensional estimate of ability was 0.0. Items were then selected using their unidimensional parameter estimates. Once an item was selected, the probability of a correct response was computed using the known two-dimensional item parameters and the preselected (θ_1, θ_2) ability in the compensatory model presented above. This probability was then compared to a randomly generated threshold value from a $U(0, 1)$ distribution to yield a correct or incorrect response. The unidimensional ability estimate was then updated. This iterative CAT process was carried out until either 20 items were administered or the selected item had an information value for the current $\hat{\theta}$ of less than .3. The CPA process was identical to the CAT simulation; however, the minimum information cutoff was set to 0.0 and the maximum test length was set to 100 items.

The calibrated $\hat{\theta}$'s were then rank ordered and divided into 20 quantiles. For each quantile, the $\bar{\theta}_1$, $\bar{\theta}_2$ centroids were calculated. Since the centroids appear to be described by a line, a least squares regression was used to provide a linear approximation to the univariate ability scale orientation in the two-dimensional plane. By comparing the scale orientation for each test format the first hypothesis could be evaluated.

Once the orientation of the univariate scale was established, the two types of testing formats were again simulated 100 times each for 37 selected points throughout the θ_1, θ_2 plane. The 37 points (Figure 3) were selected to cover the region with greatest density for the bivariate normal distribution, $N(\underset{\sim}{0}, \Sigma)$.

"Migration" vectors were then plotted to examine the second hypothesis. These vectors illustrate how the selected (θ_1, θ_2) points was mapped onto the newly oriented univariate ability scale. The mapping of the two dimensional plane onto the univariate scale was then evaluated.

To evaluate how points from the two-dimensional ability plane are mapped onto the univariate scale it is necessary to examine the response surface of the compensatory MIRT model. In Figure 4, the contour plots showing lines of equi-probability for a correct response for three equally discriminating items are shown. Three selected abilities, A(2, 0), B(0, 0), and C(0, 2) are plotted on each contour.

Item 1, shown in Figure 4a, discriminates or provides information only along θ_2. If a test was composed of items identical to Item 1, points B and C would receive the same ability estimate; however, it would be less than A.

Item 2, displayed in Figure 4b, provides information only along θ_1. In a test composed of items of this type, points A and B would receive the same ability estimate, although Point C would have a higher estimated ability.

Item 3, plotted in Figure 4c, represents an item that discriminates equally well on both θ_1 and θ_2. Points A and C would be estimated to have the same ability, by a test having items which also discriminate equally well on both dimensions. On such a test, Point B would be estimated to have a lower ability than both A and C.

Thus, by examining points on the θ_1, θ_2 plane in relationship to other points which have the same θ_1, same θ_2 or opposite (θ_1, θ_2) coordinates (i.e. (0, 3) and (3, 0) which lie on a diagonal) the degree to which each dimension is being measured can be determined.

A final simulation was used to examine the third hypothesis. Seven (θ_1, θ_2) ordered pairs representing the unidimensional abilities -3, -2, -1, 0, 1, 2, and 3 on the mapped univariate scale were computed. A CAT was simulated 100 times for each of these (θ_1, θ_2) combinations to observe if the composite of items administered at each selected ability level measured each dimension to the same degree.

Results

The centroid plots for the 20 quantiles for both test formats are shown in Figure 5. A line of best fit, representing the orientation of the univariate scale in the θ_1, θ_2 plane was obtained using a least squares regression procedure. The orientation equations are $Y = .81X - .03$ for the CPA and $Y = .25X - .03$ for the CAT.

Migration plots, illustrating where on each test format's orientation line the selected (θ_1, θ_2) abilities were mapped, are shown in Figures 6a and 6b. Several interesting comparisons can be made by examining these two plots. First the CPA ability orientation has an estimated slope that is more steep than the CAT orientation line. This would imply that those items which discriminate better along θ_1 have larger calibrated univariate discrimination parameters and would be more likely to be selected in a CAT administration. Thus it was assumed that LOGIST "oriented" the univariate scale closer to the θ_1 axis. This was confirmed by computing $r_{\hat{\theta}\theta_1}$ and $r_{\hat{\theta}\theta_2}$ which were .68 and -.50, respectively.

A second notable feature is that the CAT migration "vectors" appear to contract the ability scale whereas the complete administration (CPA) migration vector seem to expand the scale. Using the CPA format the points (3, 0) and (0, 3) are mapped onto the univariate scale in roughly the same place suggesting that the items administered at these points collectively measure both dimensions equally well. The same is true for the points (-3, 0) and (0, -3). However, in the CAT administration points (0, 3) and (0, 0) are mapped near the same univariate ability, suggesting more information in the items being presented at these points, is provided along θ_1. The same information structure occurs in the items administered at the points (0, 3) and (0, 0).

A final plot of the test information vectors was drawn for those items that were administered at the seven selected ability values along the mapped θ scale. The purpose of this plot was to see if the amount of multidimensional information was uniform throughout selected points, or if lower/higher abilities received more/less informative items along one dimension than the other. The resulting information vectors, plotted in Figure 7, also suggest that the univariate abilities received items which discriminate better along θ_1 for most of the selected abilities. In the range $-3 \leq \hat{\theta} \leq 1$, θ_1 is noticeably being measured with more accuracy.

Discussion

Several areas of caution or concern are suggested by these results. Even though an item pool when considered collectively, provides approximately the same amount of information in each direction for all points on the θ_1, θ_2 plane it does not guarantee that smaller subsets will also provide the same amount of uniform information. The orientation of the univariate ability scale appears to be a function of the informational structure of the items.

Secondly, the LOGIST calibration process tends to emphasize only some of the dimensions that may occur in the test items. Depending on how its scale is oriented in the two dimensional plane, discrimination estimates for specific items are greater or smaller. The closer the LOGIST ability scale is oriented to the item vector, the higher the discrimination parameter estimate for the item. Items with vectors that are essentially orthogonal to the line of orientation will have very small discrimination parameter estimates.

Thus the calibration orientation ultimately would determine which items will have the higher discriminations and thus be chosen more often for unidimensional CAT. While this process should suggest that items selected at the various abilities above the univariate scale should not vary too much in dimensionality, this was not the case. In the CAT simulation at the seven selected abilities (Figure 7), more information was provided along θ_1 for $-2 < \hat{\theta} < 3$. At points not on the mapped univariate scale (3, 0), (0, 3), (-3, 0), and (0, -3) administered items also provided more information along θ_1. Thus, it appears that despite using an item pool that collectively provided uniform information, different points in the ability plane did not receive what Samejima (1977) terms parallel tests.

Experiment 2

In the second part of this study, a parallel analysis was conducted using simulated data based on the characteristics of an actual item pool.

Method

Item Response Generation

The item pool used in the second part of this study was created from the ACT Assessment Math Usage test Form 26A. The test contained 40 multiple choice items covering six content areas (See Appendix A for a brief description of the content areas). Using 3,000 subject's responses from an ACT test administration, two-dimensional item parameters were estimated using the compensatory multidimensional IRT program MIRTE (Carlson, 1987). MIRTE was used to calibrate the response data using the two-dimensional IRT model given in equation 1. The calibrated multidimensional item parameters (a_1, a_2, and d) 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 set were each repeated twice to produce 16 items in all. The purpose of expanding the dataset was twofold: to increase the size of the CAT pool and to assure that each content area had the same number of items in the pool.

Three thousand pairs of(θ_1, θ_2) were then randomly selected from the bivariate normal distribution described previously. Responses for the 96-item set were then generated using the multidimensional item parameter estimates in the M2PL model. This was necessary to preserve the multidimensionality of the items. The simulated response data were then calibrated using LOGIST to obtain unidimensional item parameter estimates.

Procedure

Again two different test formats were simulated. The first test involved the complete administration (CPA) of the entire 96-item pool. The second test was a simulated adaptive test (CAT) which followed the same procedure as in Experiment 1.

Each test was simulated for 1,000 subjects randomly drawn from a bivariate normal N(0, Σ) distribution to obtain the orientation of the estimated ability scale in the θ_1, θ_2 ability plane.

To determine how the multidimensional ability scale would map onto this orientation line, each test was simulated 100 times at 37 ability points (Figure 3) selected to represent this bivariate normal distribution. A final simulation was conducted 100 times for each of the points on the mapped univariate scale representing the abilities from -3 to +3 in increments of 1.0.

Besides the migration plots for each type of test and the test information plot at selected univariate abilities, the content of items selected at each ability level was examined to see if different abilities received a different composition of items in the CAT simulation.

Results

A plot of the multidimensional IRT test information function for the 96 item pool is shown in Figure 8. Unlike the generated data set which had uniform information, the expanded Form 26A item pool provided the most information in a band where Theta 2 = 0.0 and Theta 1 spanned the range from -3 to +3. The greatest amount of information in this band is concentrated more along the second ability dimension. Very little information is provided where both abilities were very high or very low. To further explore this issue, plots of the test information functions for each of the six content areas were examined for differences. These plots are shown in Figures 9a-f.

The multidimensional information plots for the Geometry (G), the Number and Numeration Systems (NNS), the Intermediate Algebra (IA) and the Algebraic and Arithmetic Operations (AAO) items all have larger information vectors along the second ability dimension, albeit at different places in the θ_1, θ_2 ability plane. The Arithmetic and Algebraic Reasoning (AAR) items are more discriminating along the first ability dimension.

Upon examining the content of these item types, it becomes clear that the first dimension might correspond to a verbal reasoning ability because all of the AAR items are "story" or word problems. The other content areas, IA, G, NNS, and AAO all involve some form of numerical computation. This hypothesis was confirmed by examining the advanced topic (AT) information plot. On the original 40 item test there were only two AT items. One of the items was a story problem, the other an algebraic manipulation problem. The information vectors for AT tend to measure best along θ_2 for the horizontal band and best along θ_1 in the vertical band. It is believed that the horizontal band illustrates the information provided by the numerical computation items and the vertical band the information provided by the verbal reasoning item.

To verify that the variety of multidimensional test information surfaces for the six contents would have equally dissimilar univariate test information functions, the test information plot using the LOGIST item parameter estimates for each content area and the total test is shown in Figure 10. The AAO content area provided the most overall information, while NNS provided the least. The six univariate test information functions provided a striking contrast to their multidimensional counterparts. Compared to the wide variation portrayed in Figure 8, the univariate plots show a higher degree of similarity.

The plots of the centroids for the CAT and CPA administrations are shown in Figure 11. The curvature of the centroids, suggests a confounding of difficulty and dimensionality. Each test's centroids were fit using a hyperbolic function:

$$\theta_2 = \frac{.3 - 2\theta_1}{\theta_1 - .9} \text{ for the CAT centroids,}$$

$$\text{and } \theta_2 = \frac{.3 - 2.5\theta_1}{\theta_1 - 2.75} \text{ for the CPA centroids.}$$

The migration plots, showing how the two different test administration formats mapped the selected points onto the orientation curve, are illustrated in Figure 12a and b. Although the orientation of the univariate scale is more curved for the CAT, the mapping of the 37 selected abilities appears to be somewhat similar. The contraction/expansion difference between the CAT and CPA mappings that existed with the generated item pool in Experiment 1 does not appear. Most of the abilities are mapped orthogonally onto the univariate ability scale.

In the CPA, the ability points (3, 0) and (0, 3) are mapped into about the same univariate ability. This is also the case for the points (-3, 0) and (0 -3). This would suggest that these four points receive items in CPA, which collectively discriminate θ_1 and θ_2 equally well.

In the CAT, abilities (-3, 0) and (0, -3) are mapped close to the same univariate ability. However, the points (0, 0) and 0, 3) are mapped closer to each other than (3, 0) and (0, 3). This would indicate that points in the first quadrant receive items that discriminate better along θ_2, while the rest of the points receive items which discriminate equally well between both dimensions.

The CAT information function for the seven selected ability points along the mapped univariate scale is shown in Figure 13. Unlike the results of Experiment 1, items selected at each ability level appear to be measuring the two ability dimensions equally well.

The results of the content composition at the seven abilities were also computed. Table 1 demonstrates the shift in item content for each of the ability levels. No NNS items were selected at any of the seven ability values. This may be the result of their low univariate discrimination values which is partly due to the LOGIST scale orientation in the θ_1, θ_2 plane. This is further verified by the univariate test information plot (Figure 10) which shows the NNS items as being least discriminating. Only the IA items appear to be represented in about the same percentage across the $\hat{\theta}$ scale. Fifty percent of the items administered at $\hat{\theta} = -3.0$ were AAO items, while at $\hat{\theta} = 3.0$ only 16% items were AAO. No AT items were administered below a $\hat{\theta}$ value of 0.0, yet over a fourth of the items administered at $\hat{\theta} = 3.0$ were AT.

A check to see if the use of the multidimensional item parameters affected the estimates in any way was conducted. A second CAT administration was simulated at the same ability points using only the univariate item parameter estimates to calculate the probability of a correct response. The difference in item content is even more pronounced. At $\hat{\theta} = -3.0$, 88% of the items administered were IA items. At $\hat{\theta} = 3.0$, 89% of the items administered were AT items. Again no NNS items were selected.

General Discussion

The results of this study would strongly suggest that the concerns over different abilities receiving different content in adaptive testing are valid. In Experiment 2, the multidimensional item vectors computed at the seven selected $\hat{\theta}$ points along the mapped univariate scale appeared to be measuring equal composites of θ_1 and θ_2 (Figure 13). However, the percentages of the six content areas administered throughout the univariate ability scale did differ noticeably. It would be interesting to see how dramatic the shift in $\hat{\theta}$ would be if the CAT was administered in such a way as to control the number of items from each content area. Because each CAT test may differ in length, the problem would not easily be solved.

The orientation of the univariate ability scale in the two-dimensional ability plane appears to be a function of multidimensional composition of the items administered in a test. The item selection process of a CAT administration tends to make the univariate scale orientation similar to that obtained for a univariate calibration. This occurs because items which have θ_1, θ_2 composites closer to the univariate calibration orientation will have higher estimated discrimination values.

Multidimensional information vector plots appear to be quite helpful in revealing suspected differences between different content areas. While unidimensional information functions may display a high degree of similarity, multidimensional information plots could be used to help identify the necessary component skills required to answer various item types. Such information could become an important ingredient in the test development process.

This study also graphically illustrated how different two-dimensional abilities are mapped onto a univariate scale which is oriented in the θ_1, θ_2 ability plane. For both generated data and the quasi-real data there appeared to be differences in the degree to which each ability dimension was measured for different points in the θ_1, θ_2 there plane. That is, strictly parallel tests would not be administered at all (θ_1, θ_2) points in the ability plane for either data set. Perhaps this could only occur for a truly unidimensional item pool.

This study suggests that more work needs to be conducted to understand the purifying process that use of unidimensional calibration of multidimensional data has on the ability estimates obtained from a CAT administration. The richness of different item contents may have to be filtered out to meet the model requirements of univariate CAT.

REFERENCES

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

Reckase, M. D. (April, 1986). The discriminating power of items that measure more than one dimension. Paper presented at the AERA Annual Meeting, San Francisco.

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

Reckase, M. D., Carlson, J. E., Ackerman, T. A., & Spray, J. A. (June, 1986). The interpretation of unidimensional IRT parameters when estimated from multidimensional data. Paper presented at the Psychometric Society Annual Meeting, Toronto.

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). British Columbia: Educational Research Institute of British Columbia.

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

TABLE 1

The Percent of Each Content Area Sampled in the Multidimensional CAT Simulation at Selected Ability Levels

	Ability Level						
Content	-3.0	-2.0	-1.0	0.0	1.0	2.0	3.0
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

TABLE 2

The Percent of each Content Area Sampled in the Univariate CAT Simulation at Selected Ability Levels

	Ability Level						
Content	-3.0	-2.0	-1.0	0.0	1.0	2.0	3.0
AAR	.07	.06	.14	.14	.13	.11	.11
NNS	.00	.00	.00	.00	.00	.00	.00
IA	.88	.41	.26	.27	.28	.31	.00
G	.02	.12	.23	.21	.22	.20	.00
AAO	.03	.41	.36	.37	.36	.33	.00
AT	.00	.00	.01	.01	.01	.05	.89

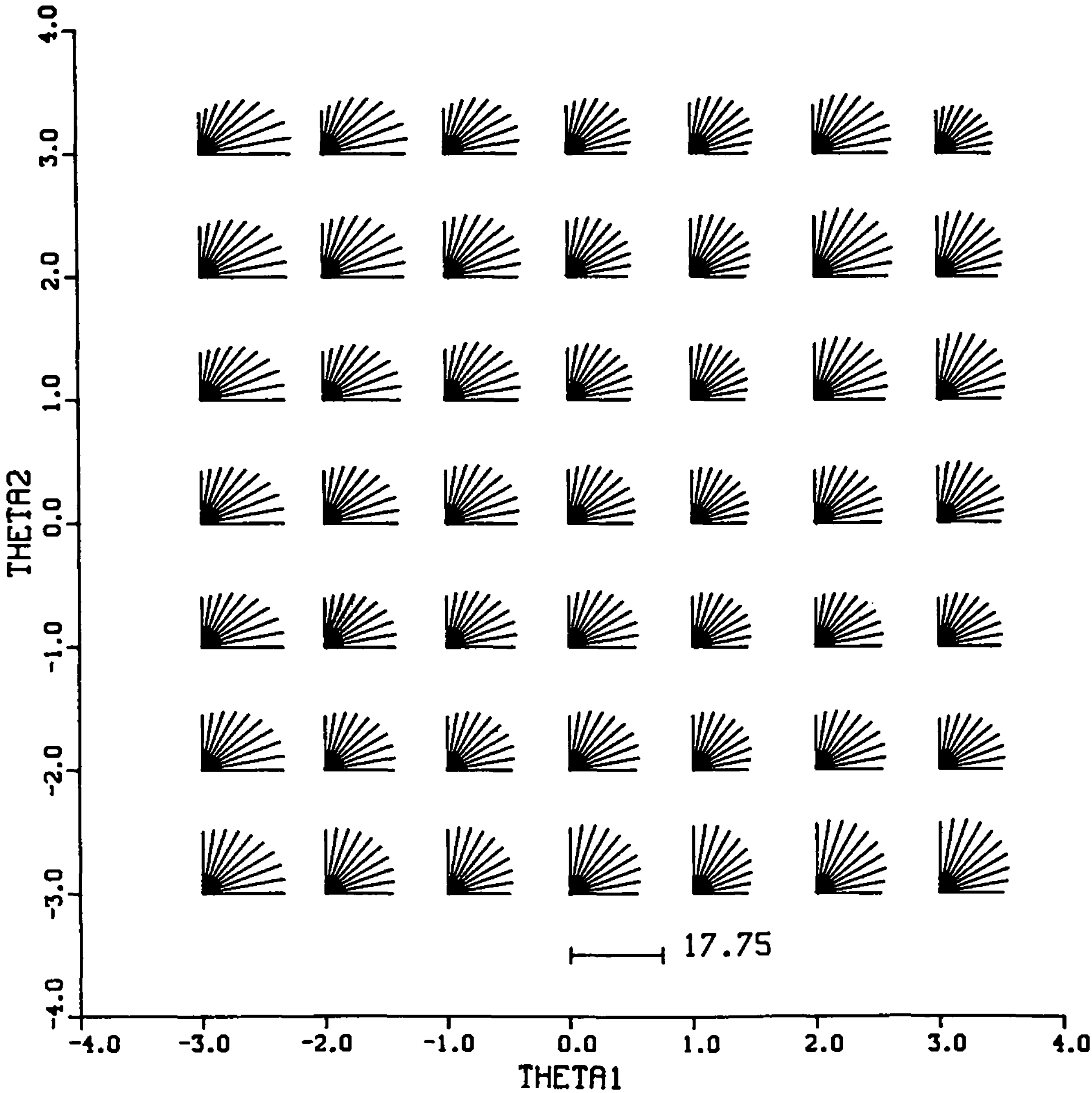


Figure 1. Multidimensional Test Information Vectors for the Generated Item Parameters.

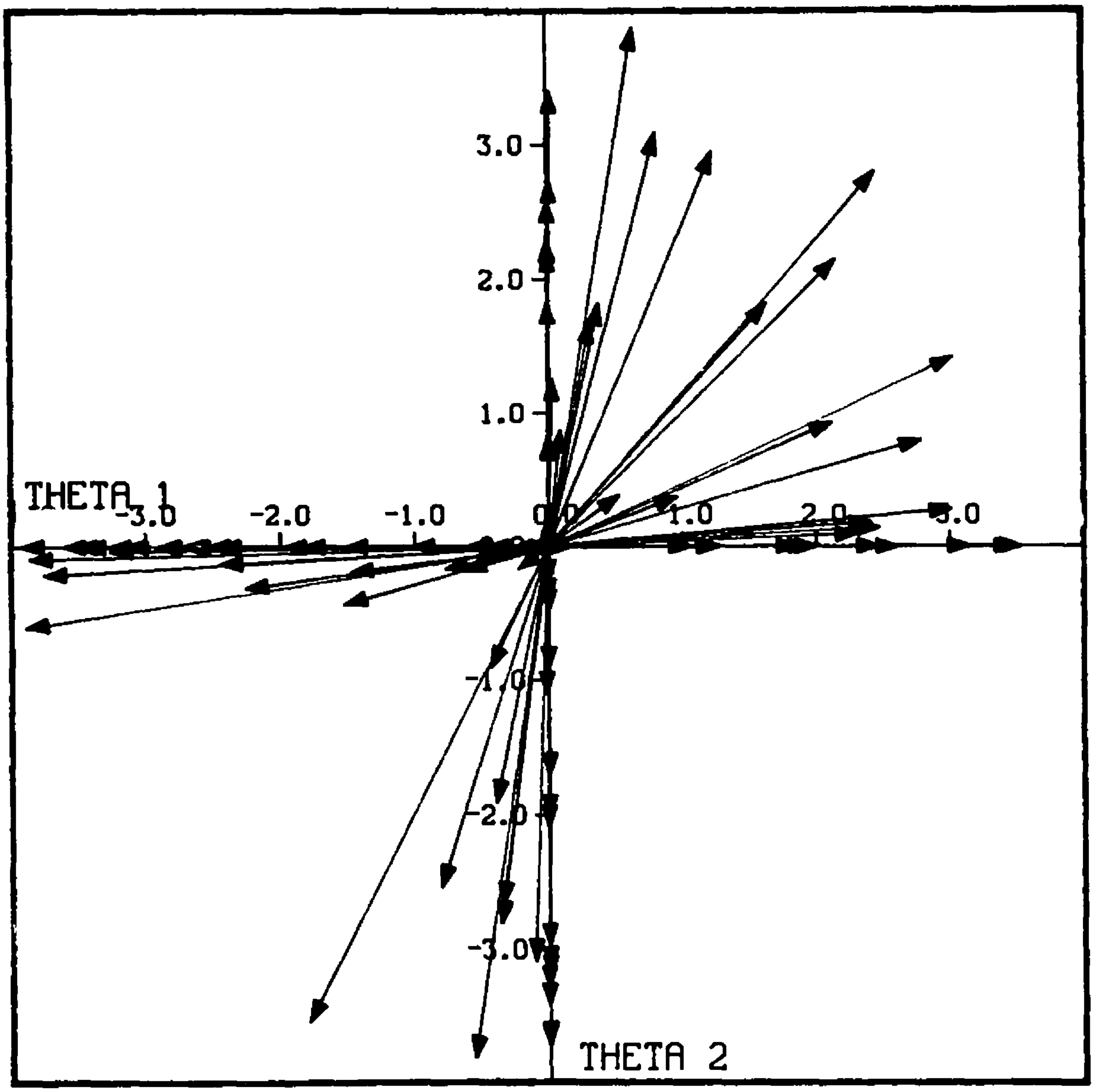


Figure 2. Item Vectors Corresponding to the 100 Generated Item Parameters.

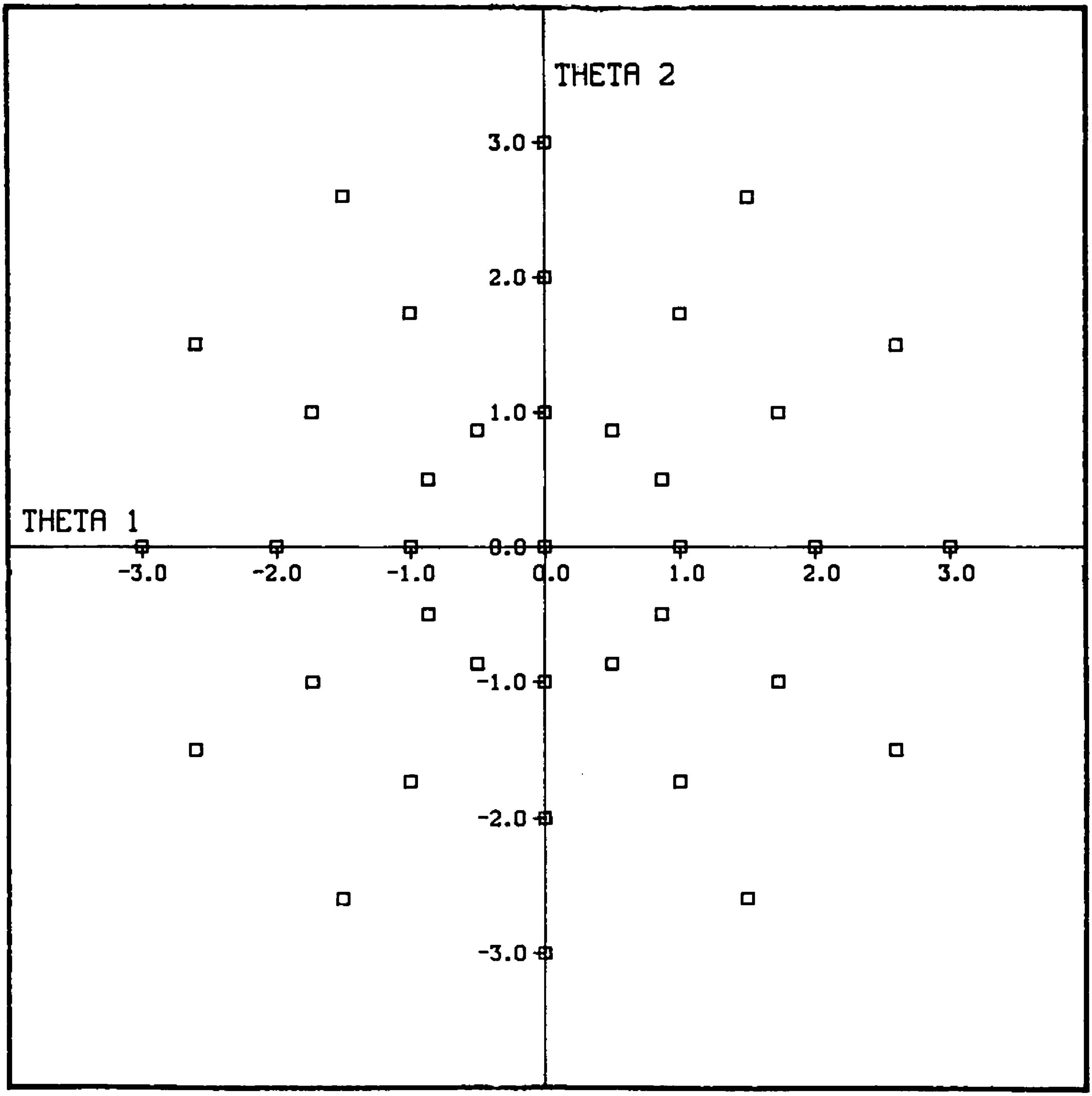
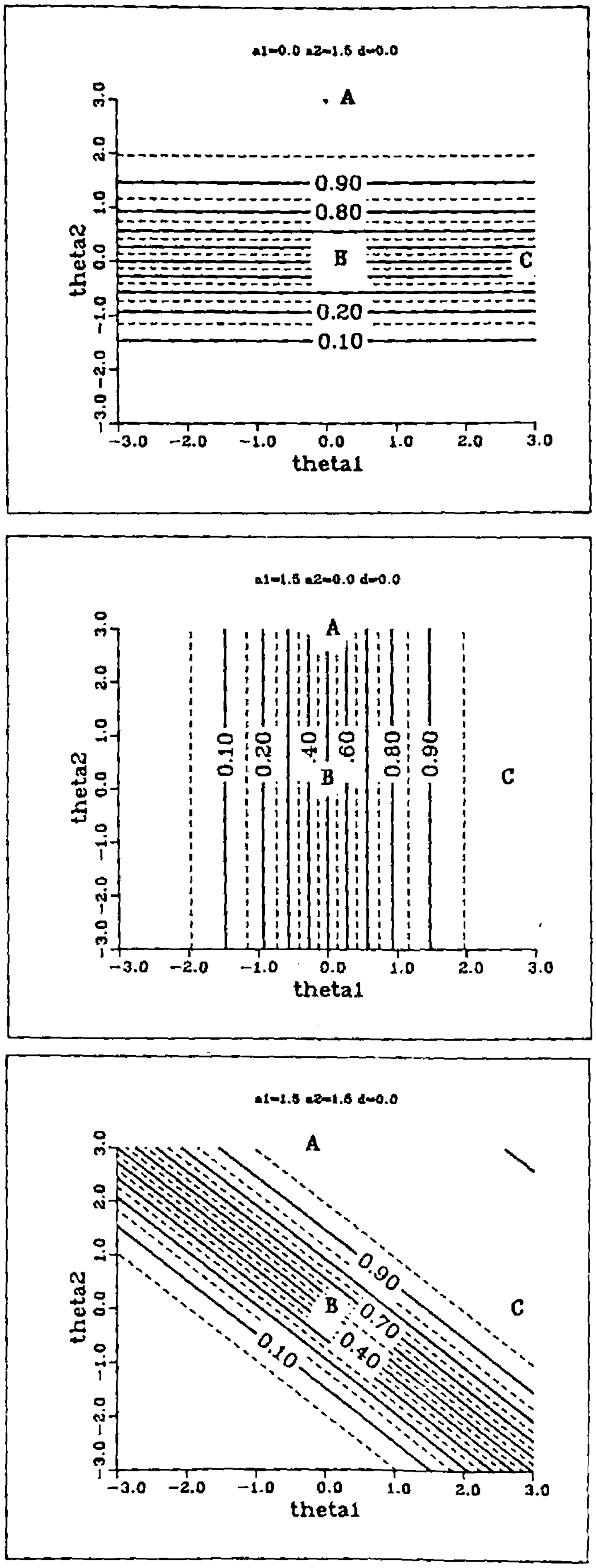


Figure 3. Points in the Two-Dimensional Ability Space Used to Represent the Bivariate Normal Distribution.



4a

4b

4c

Figure 4. Contour Plots of Three Compensatory Items which Differentiate along θ_2 only (4a), along θ_1 only (4b), and equally on θ_1 and θ_2 (4c).

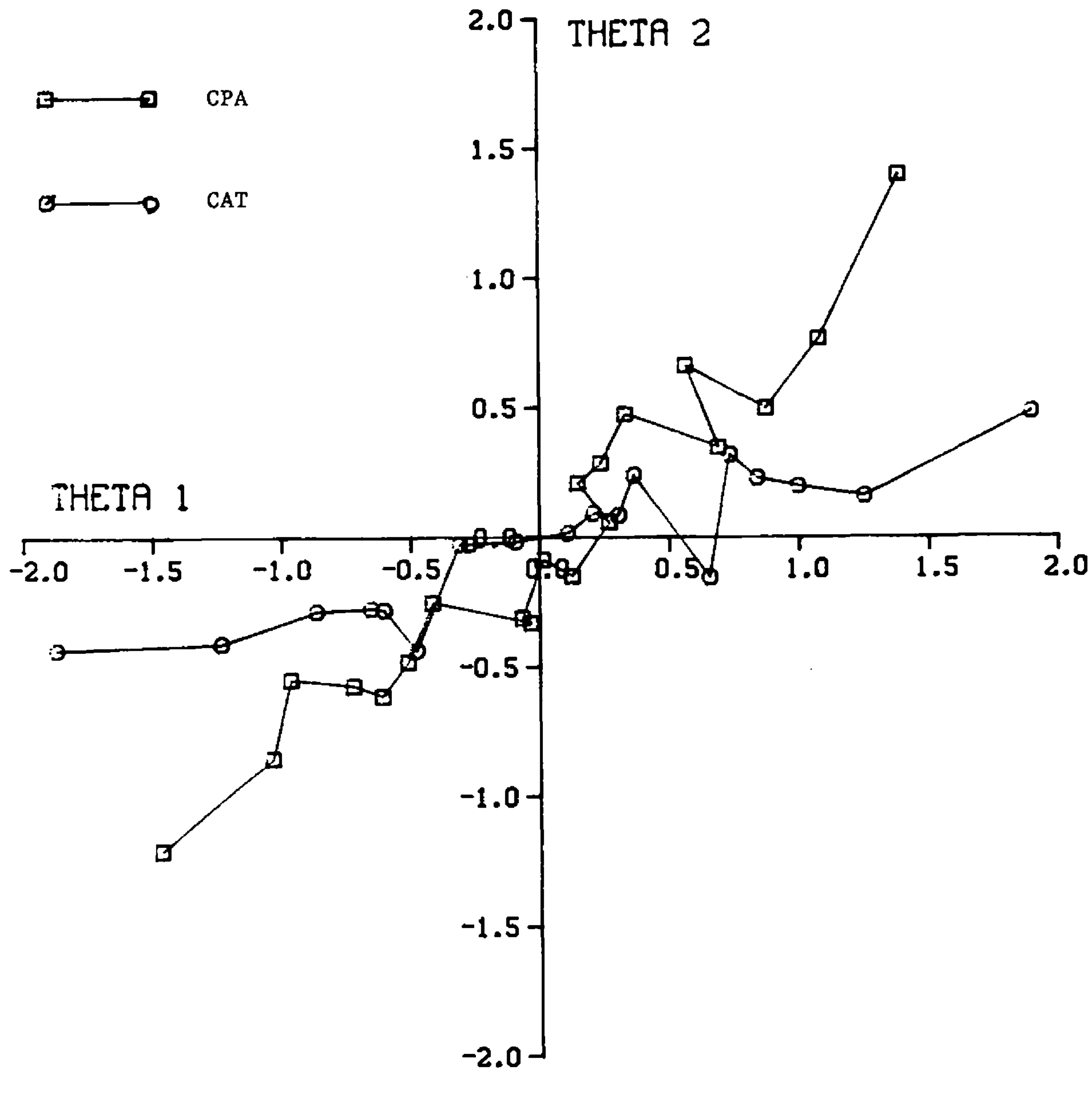
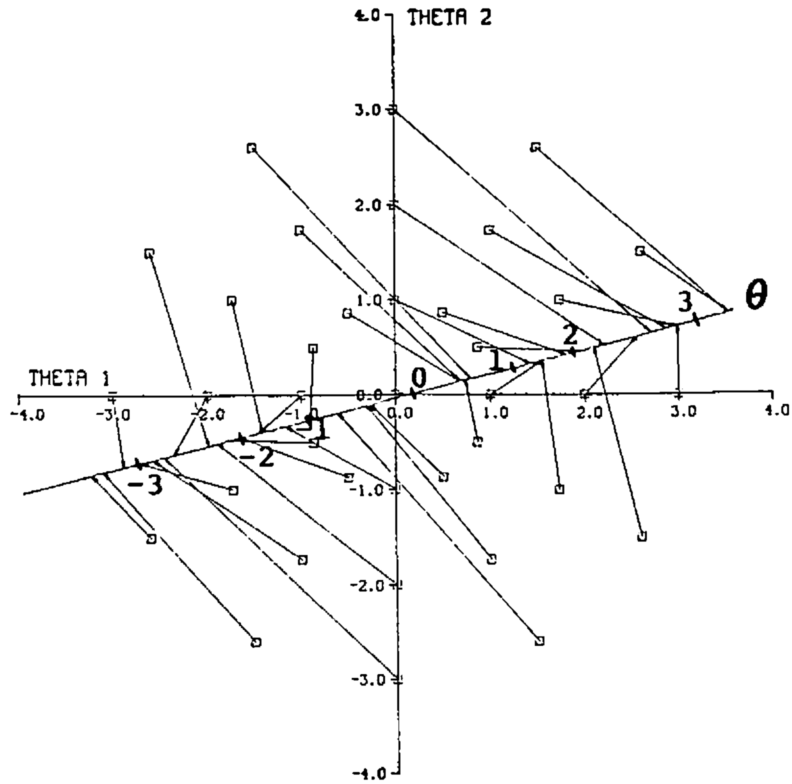
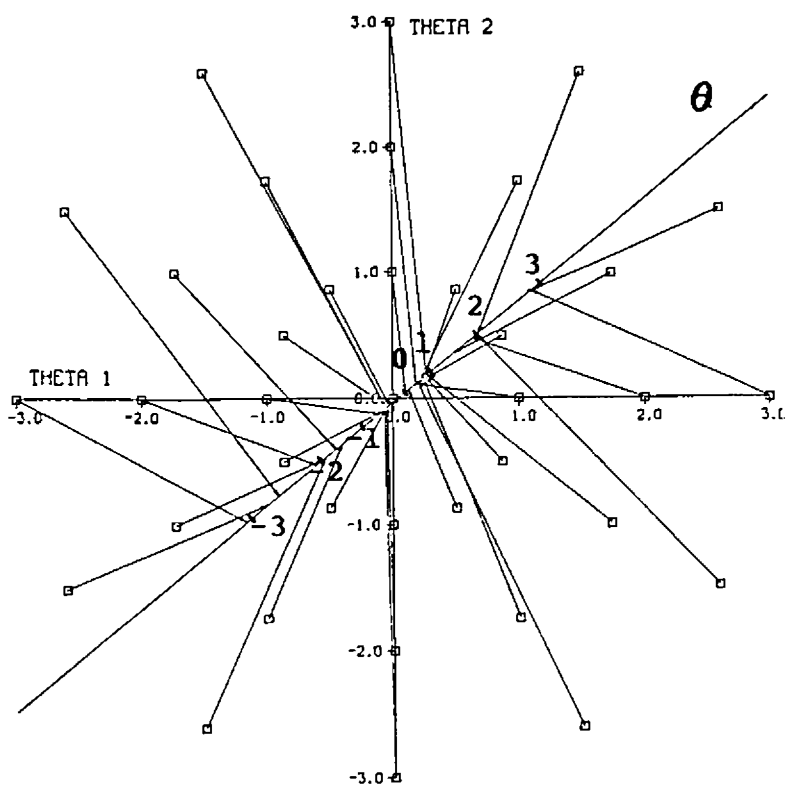


Figure 5. The Centroids of the Observed 20 Quantiles for the Simulated CAT and CPA test administrations.



6a



6b

Figure 6. Migration Plots for the Simulated CPA (6a) and CAT (6b) Test Administrations.

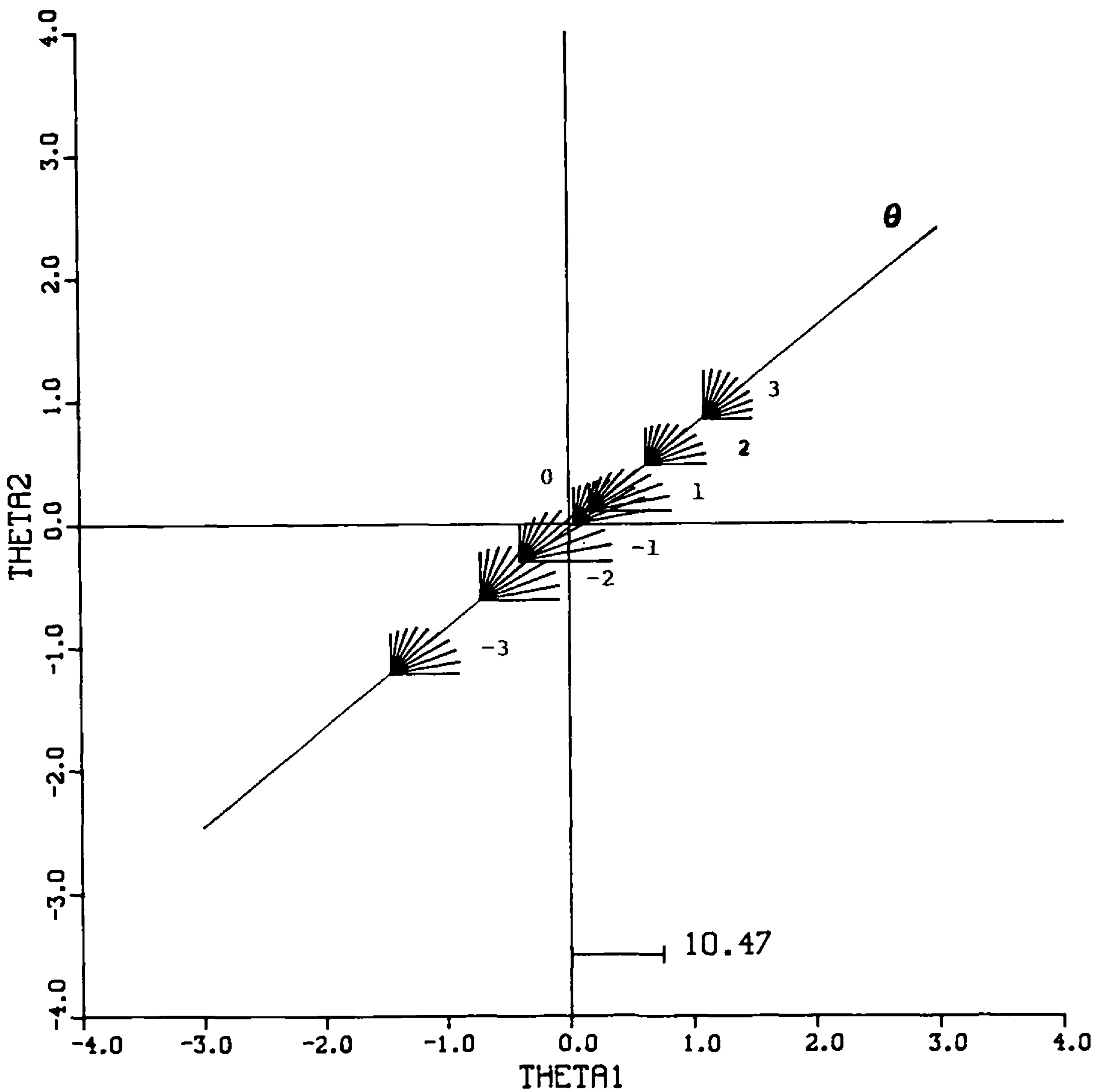


Figure 7. Multidimensional Test Information Vectors of the Generated Item Parameters Administered in the CAT at Seven Selected Abilities Along the Mapped Univariate Ability Scale.

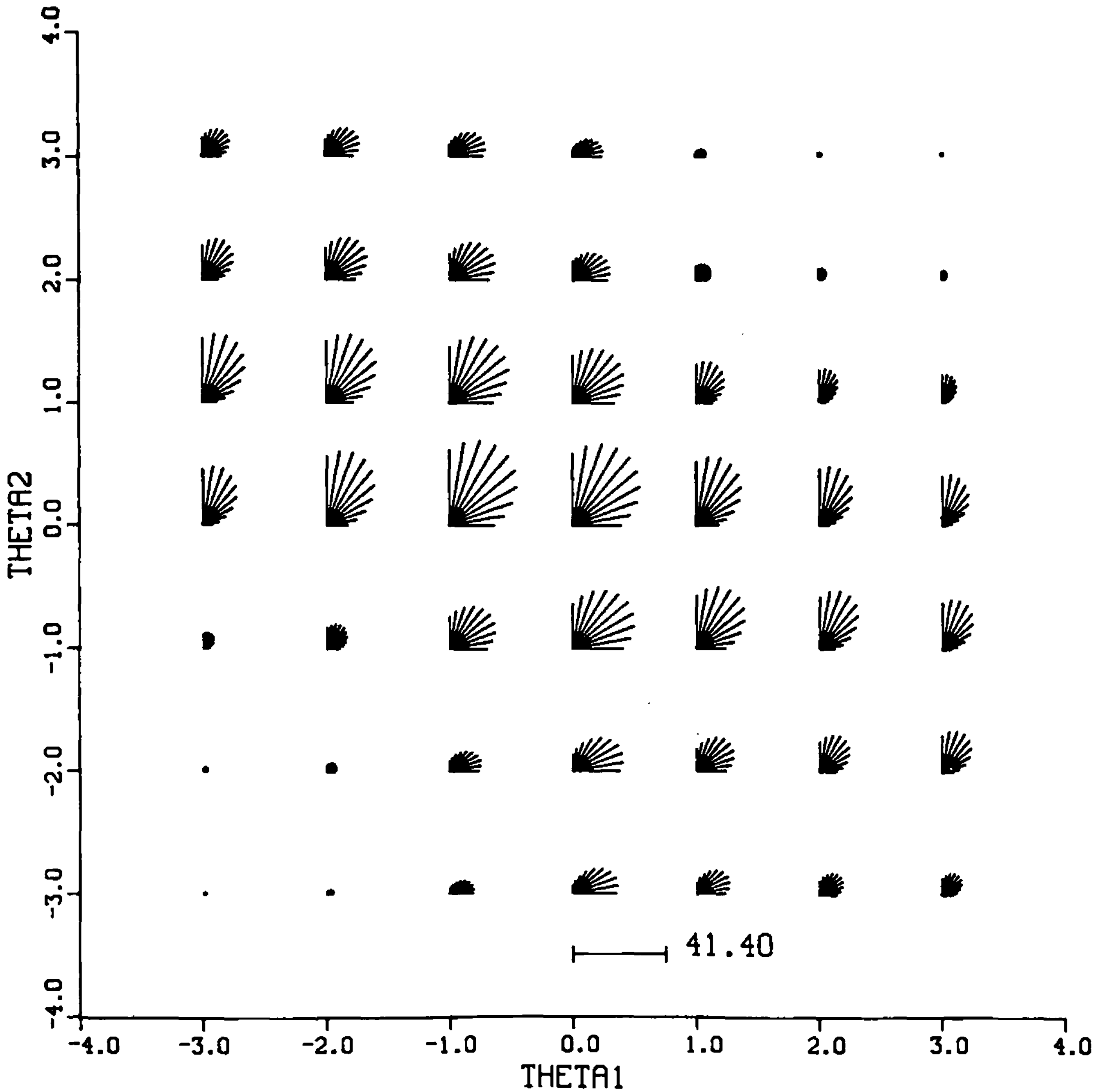


Figure 8. Multidimensional Test Information Vectors for Item Parameters from the Expanded Form 26A.

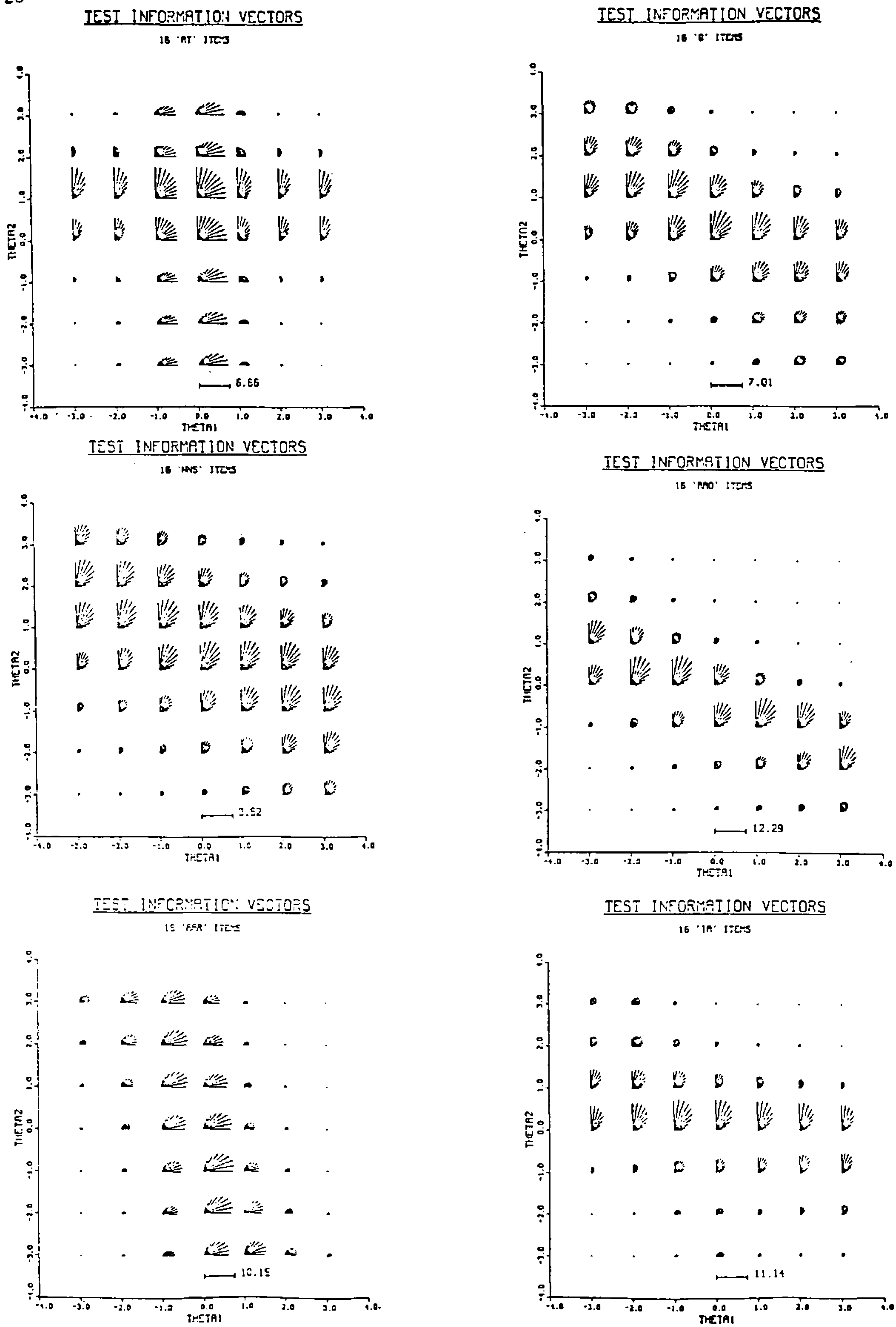


Figure 9. Multidimensional Test Information Vectors for Each of the Six Content Areas.

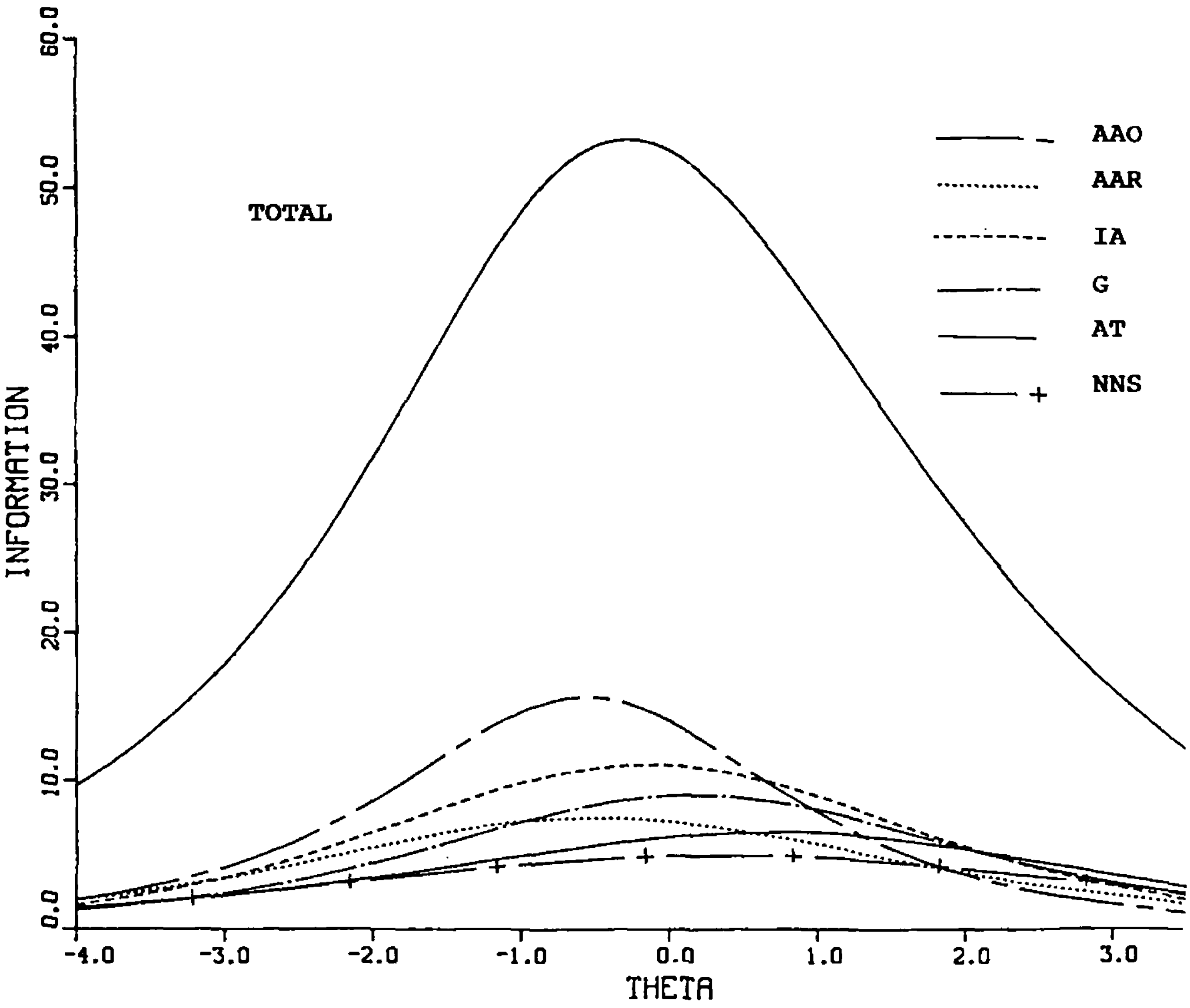


Figure 10. Unidimensional Test Information Functions for the Total Pool and Each Content Area Using the Calibrated Item Parameters.

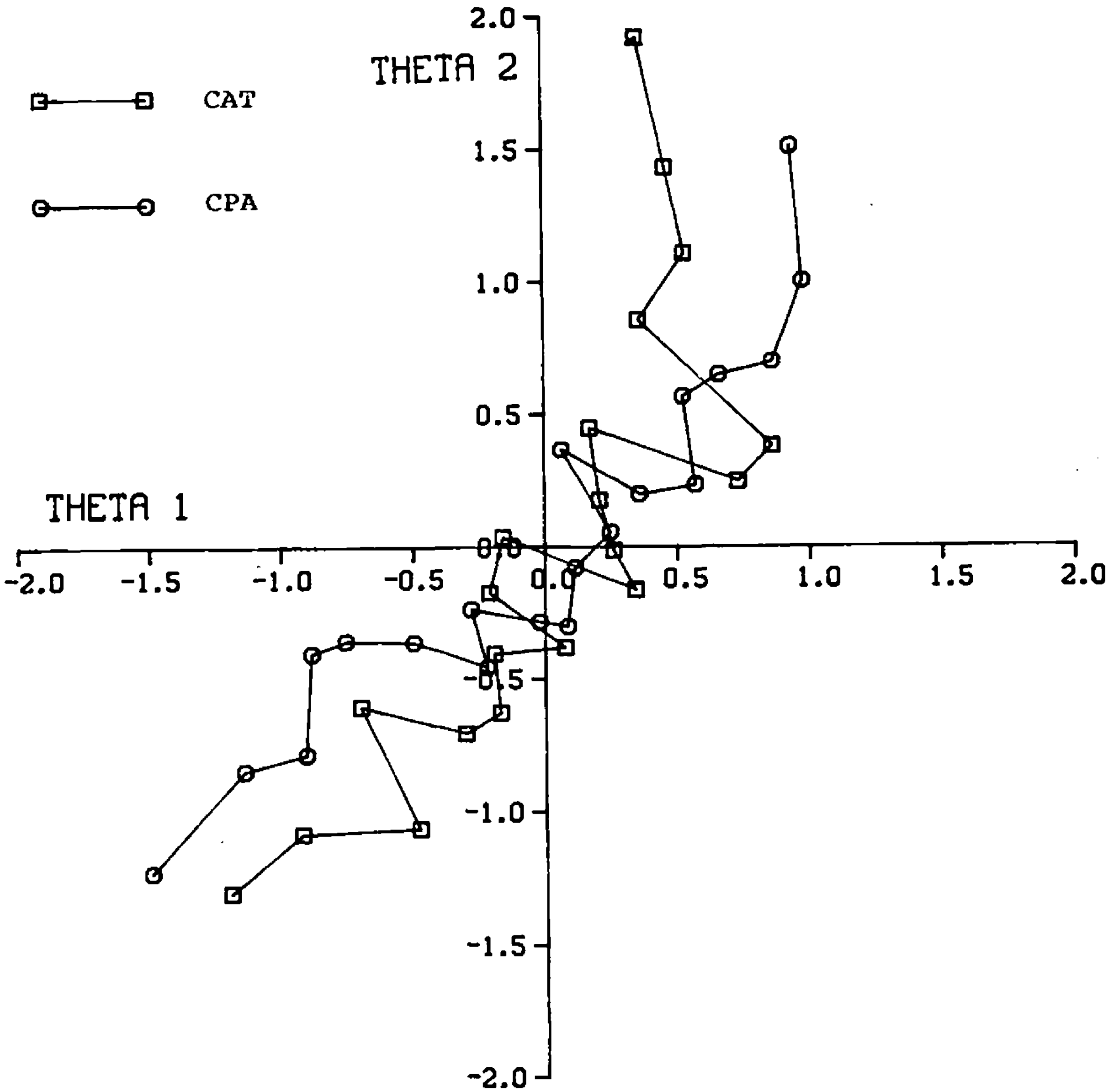
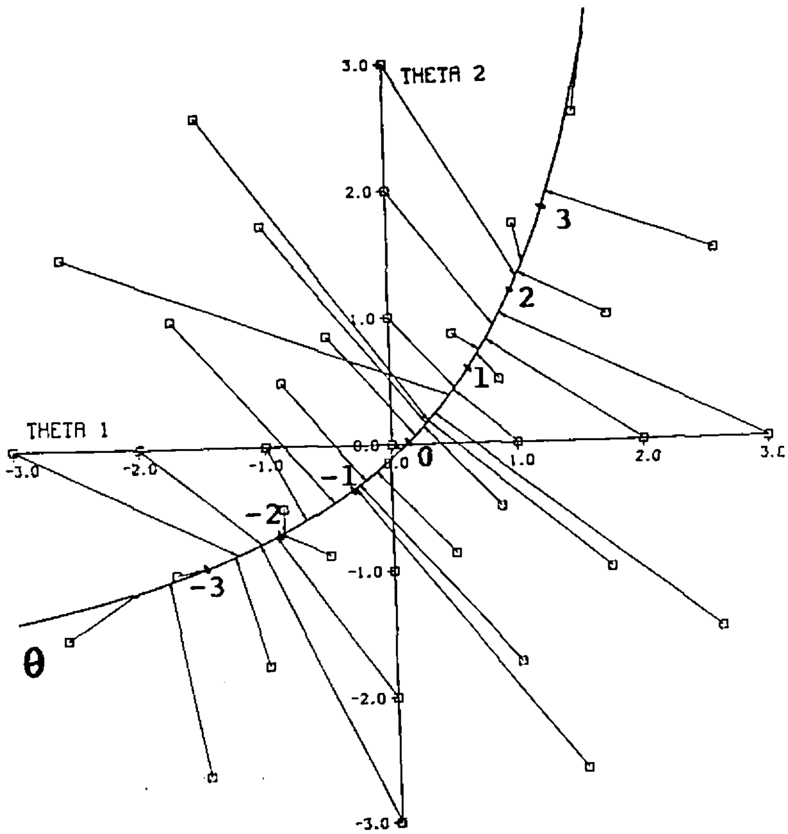
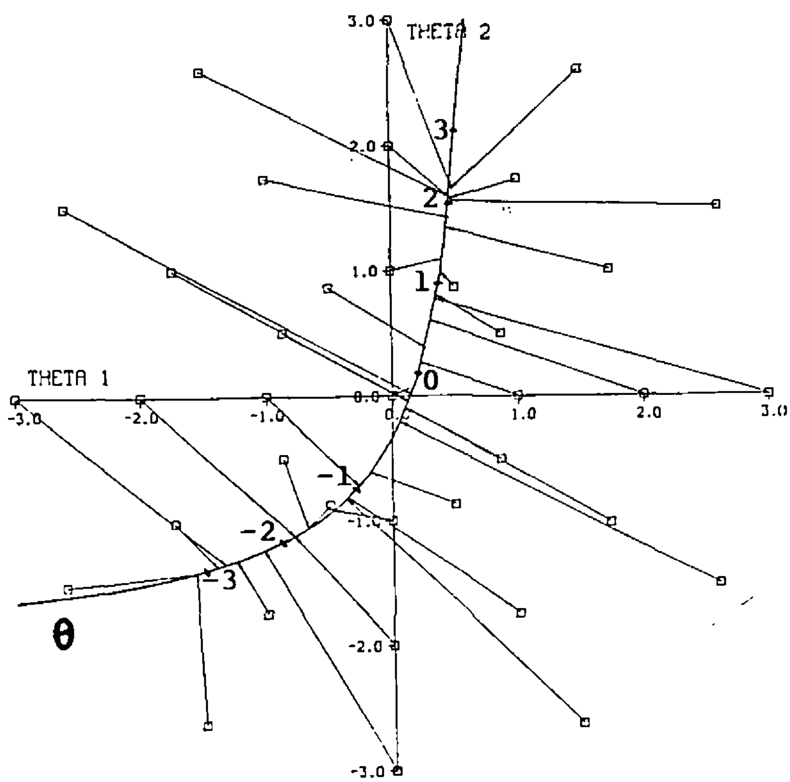


Figure 11. The Centroids of the Observed 20 Quantiles for the Simulated CAT and CPA Test Administrations Using the Item Parameters from the Expanded Form 26A.



12a



12b

Figure 12. Migration Plots for the Simulated CPA (12a) and CAT (12b) Test Administrations Using the Expanded Form 26A Item Pool.

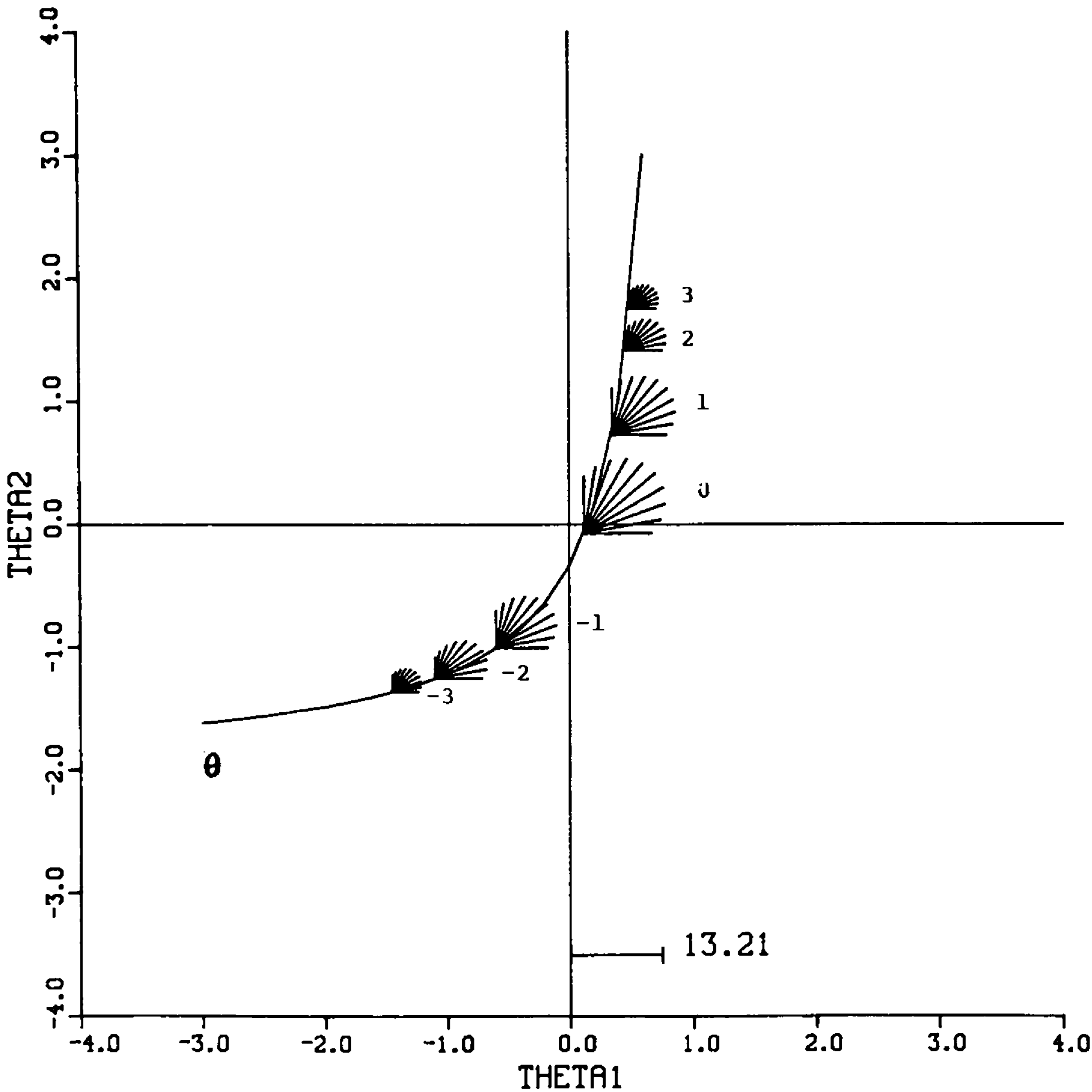


Figure 13. Multidimensional Test Information Vectors of the Items from Form 26A Administered in the CAT at Seven Selected Abilities Along the Mapped Univariate Ability Scale.

Appendix A

ACT MATHEMATICS USAGE TEST

Description of the test. The Mathematics Usage Test is a 40-item, 50-minute test that measures the students' mathematical reasoning ability. It emphasizes the solution of practical quantitative problems that are encountered in many postsecondary curricula and includes a sampling of mathematical techniques covered in high school courses. The test emphasizes quantitative reasoning, rather than memorization of formulas, knowledge of techniques, or computational skill. Each item in the test poses a question with five alternative answers, the last of which may be "None of the above."

Content of the test. In general, the mathematical skills required for the test involve proficiencies emphasized in high school plane geometry and first- and second-year algebra. Six types of content are included in the test. These categories and the approximate proportion of the test devoted to each are given below.

Mathematics Content Area		Proportion of Test	Number of Items
a. Arithmetic and Algebraic Operations		.10	4
b. Arithmetic and Algebraic Reasoning		.35	14
c. Geometry		.20	8
d. Intermediate Algebra		.20	8
e. Number and Numeration Concepts		.10	4
f. Advanced Topics		.05	2
	Total	1.00	40

a. *Arithmetic and Algebraic Operations.* The items in this category explicitly describe operations to be performed by the student. The operations include manipulating and simplifying expressions containing arithmetic or algebraic fractions, performing basic operations in polynomials, solving linear equations in one unknown, and performing operations on signed numbers.

b. *Arithmetic and Algebraic Reasoning.* These word problems present practical situations in which algebraic and/or arithmetic reasoning is required. The problems require the student to interpret the question and either to solve the problem or to find an approach to its solution.

c. *Geometry.* The items in this category cover such topics as measurement of lines and plane surfaces, properties of polygons, the Pythagorean theorem, and relationships involving circles. Both formal and applied problems are included.

d. *Intermediate Algebra.* The items in this category cover such topics as dependence and variation of quantities related by specific formulas, arithmetic and geometric series, simultaneous equations, inequalities, exponents, radicals, graphs of equations, and quadratic equations.

e. *Number and Numeration Concepts.* The items in this category cover such topics as rational and irrational numbers, set properties and operations, scientific notation, prime and composite numbers, numeration systems with bases other than 10, and absolute value.

f. *Advanced Topics.* The items in this category cover such topics as trigonometric functions, permutations and combinations, probability, statistics, and logic. Only simple applications of the skills implied by these topics are tested.