DOCUMENT RESUME

ED 222 497 TM 820 241

AUTHOR Sinnott, Loraine T.
TITLE The Identification of Biased Items.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-82-3
PUB DATE Jan 82
NOTE 30p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Entrance Examinations; *Difficulty Level; Higher Education; *Item Analysis; Research Methodology; Sex Differences; *Test Bias; *Test Items; *Test Reliability
IDENTIFIERS Graduate Management Admission Test; Multidimensional Approach

ABSTRACT

A standard method for exploring item bias is the intergroup comparison of item difficulties. This paper describes a refinement and generalization of this technique. In contrast to prior approaches, the proposed method deletes outlying items from the formulation of a criterion for identifying items as deviant. It also extends the mathematical framework of item difficulty comparisons to allow the simultaneous analysis of any number of groups. As an example, the proposed method is applied to a set of quantitative items selected from a business school admission test. (Author)

RR-82-3

RESEARCH REPORT

THE IDENTIFICATION OF BIASED ITEMS

Loraine T. Sinnott

January 1982

Educational Testing Service
Princeton, New Jersey

The Identification of Biased Items

Loraine T. Sinnott

Educational Testing Service
Princeton, New Jersey
January 1982

Abstract

A standard method for exploring item bias is the intergroup comparison of item difficulties. This paper describes a refinement and generalization of this technique. In contrast to prior approaches, the proposed method deletes outlying items from the formulation of a criterion for identifying items as deviant. It also extends the mathematical framework of item difficulty comparisons to allow the simultaneous analysis of any number of groups. As an example, the proposed method is applied to a set of quantitative items selected from a business school admission test.

The Identification of Biased Items

Introduction

The study of item bias is concerned with the internal consistency of a test. An attempt is made to identify items that behave differently from other items presumed to be measuring the same ability. Implicit in research on item bias is group comparison; items are biased in favor of or against one group of test takers relative to another. Numerous techniques have emerged to investigate item bias (for a review, see Rudner, Getson, & Knight, 1980), but among the most commonly used is the intergroup comparison of item difficulties (see, e.g., Angoff & Ford, 1973; Donlon, Hicks, & Wallmark, 1980). In this technique, an item's difficulty is taken to be the z score associated with the proportion of test takers responding correctly to the item. For a set of items, the difficulties for one group are plotted against those for a second group. When the items are more or less homogeneous in the ability they measure, a line is suggested by the resulting points. This follows, since items that are more difficult for one group will be more difficult for the other group, and the easier items for one group will also be the easier items for the second group.

A line of best fit is calculated for the plotted points. Items far removed from the line behave unexpectedly relative to most other items. They could be described as more difficult for one of the groups than would have been predicted by the relative performance of the two groups on other items. One presumes that such a deviant item is sensitive to factors to which most other items are insensitive or less sensitive. By introducing additional conditions for the successful completion of the item, these factors interfere with the item's expected relative difficulty.

Thus, the comparison of item difficulties distinguishes items that behave unusually relative to the behavior of most other items, where distance from the line of best fit is the measure of unexpected behavior. This line itself represents an ideal, the relationship of a set of items that are homogeneous in what they measure.

Error enters into the approximation of the ideal line when items sensitive to extraneous factors are included in the best-fit calculations. A better approximation is achieved if the calculations include only the more homogeneous items. One can lessen the influence of items sensitive to other factors by removing from the calculations items far removed from the line suggested by the mainstream of points. A recent report presented an algorithm for doing this (Sinnott, 1980). Basically, the algorithm successively removes subsets of items, stopping when a line is found in which only those points within a specified distance participate in its derivation.

The comparison of item difficulties has thus far been restricted to two groups. In this paper procedures are outlined that allow the simultaneous comparison of any number of groups. The procedures incorporate the algorithm described above, removing from the best-fit calculations those items most likely to be sensitive to extraneous factors. After the procedures are presented, their application will be illustrated. First, though, the mathematical foundation for the procedures is presented.

Mathematical Background

In this section the line of best fit is derived for a set of points in n-space. The line sought is that which minimizes the squared distances of the points from the line. The discussion is adapted from arguments presented by Pearson (1901).

Let $\Delta_{\alpha j}$ be the difficulty of item α for group j, and S be the set of vectors $\underset{\sim}{\Delta}_\alpha = (\Delta_{\alpha 1},\ldots,\Delta_{\alpha n})$, $\alpha=1,\ldots,\gamma$, where γ is the number of items and n the number of groups. For a given group j, a mean and variance are defined by

$$M_j = \sum_{\alpha=1}^{\gamma} \Delta_{\alpha j} / \gamma \quad \text{and} \tag{1}$$

$$s_j^2 = \sum_{\alpha=1}^{\gamma} (\Delta_{\alpha j} - M_j)^2 / (\gamma - 1) \tag{2}$$

For two groups, j and k , a correlation coefficient is given by

$$r_{jk} = \sum_{\alpha=1}^{\gamma} (\Delta_{\alpha j} - M_j)(\Delta_{\alpha k} - M_k) / (\gamma - 1)s_j s_k \tag{3}$$

A line, L, in n-space can be written in the form $\underset{\sim}{x}=\underset{\sim}{x}'+t\underset{\sim}{u}$, where $\underset{\sim}{x}'$ is a vector lying on the line, t varies over the real numbers, and $\underset{\sim}{u}$ is a unit vector parallel to the line. Our goal is to find $\underset{\sim}{x}'$ and $\underset{\sim}{u}$ for the line that best fits S, expressing $\underset{\sim}{x}'$ and $\underset{\sim}{u}$ solely in terms of the statistics M_j, s_j, and r_{jk}.

Let $p(\alpha)$ be the perpendicular distance of $\underset{\sim}{\Delta}_\alpha$ from L. As will now be shown, $p(\alpha)$ can be expressed in terms of $\underset{\sim}{x}'$, $\underset{\sim}{u}$, and $\underset{\sim}{\Delta}_\alpha$. For a given $\underset{\sim}{\Delta}_\alpha$, let t' be chosen such that the vector

$$\underset{\sim}{\Delta}_\alpha - (\underset{\sim}{x}' + t'\underset{\sim}{u}) \tag{4}$$

is perpendicular to $\underset{\sim}{u}$. The length of (4) is the perpendicular distance of $\underset{\sim}{\Delta}_\alpha$ from L, or p(α). Thus

$$p^2(\alpha) = (\underset{\sim}{\Delta}_\alpha - (\underset{\sim}{x}' + t'\underset{\sim}{u}))\cdot(\underset{\sim}{\Delta}_\alpha - (\underset{\sim}{x}' + t'\underset{\sim}{u})).$$

Since $\underset{\sim}{u}$ is perpendicular to (4),

$$p^2(\alpha) = (\underset{\sim}{\Delta}_\alpha - (\underset{\sim}{x}' + t'\underset{\sim}{u}))\cdot(\underset{\sim}{\Delta}_\alpha - \underset{\sim}{x}'). \qquad (5)$$

The perpendicularity of $\underset{\sim}{u}$ and (4) is further exploited to find an expression for t′. Since $\underset{\sim}{u}\cdot(\underset{\sim}{\Delta}_\alpha - (\underset{\sim}{x}' + t'\underset{\sim}{u})) = 0$, $t' = \underset{\sim}{u}\cdot(\underset{\sim}{\Delta}_\alpha - \underset{\sim}{x}')$. Substituting this expression for t′ into (5) yields

$$p^2(\alpha) = (\underset{\sim}{\Delta}_\alpha - \underset{\sim}{x}')\cdot(\underset{\sim}{\Delta}_\alpha - \underset{\sim}{x}') - (\underset{\sim}{u}\cdot(\underset{\sim}{\Delta}_\alpha - \underset{\sim}{x}'))^2. \qquad (6)$$

Equation 6 expresses p(α) in terms of $\underset{\sim}{x}'$, $\underset{\sim}{u}$, and $\underset{\sim}{\Delta}_\alpha$. In terms of p(α), the line desired is that which minimizes $\sum_{\alpha=1}^{\gamma} p^2(\alpha)$. This sum can be written

$$\sum_{\alpha=1}^{\gamma} p^2(\alpha) = \sum_{\alpha=1}^{\gamma} (\underset{\sim}{\Delta}_\alpha - \underset{\sim}{x}')\cdot(\underset{\sim}{\Delta}_\alpha - \underset{\sim}{x}') - (\underset{\sim}{u}\cdot(\underset{\sim}{\Delta}_\alpha - \underset{\sim}{x}'))^2$$

$$= \sum_{\alpha=1}^{\gamma} \left(\sum_{j=1}^{n} (\Delta_{\alpha j} - x_j')^2 - \left(\sum_{j=1}^{n} u_j(\Delta_{\alpha j} - x_j')\right)^2\right). \qquad (7)$$

The Lagrange multiplier method can be applied to minimize (7) subject to the constraint that $\sum_{j=1}^{n} u_j^2 = 1$. This will lead to expressions for $\underset{\sim}{x}'$ and $\underset{\sim}{u}$ in terms of the $\underset{\sim}{\Delta}_\alpha$ for the line that best fits S. The Lagrange formula is

$$\sum_{\alpha=1}^{\gamma} \left(\sum_{j=1}^{n} (\Delta_{\alpha j} - x_j')^2 - \left(\sum_{j=1}^{n} u_j(\Delta_{\alpha j} - x_j')\right)^2\right) + \lambda\left(\sum_{j=1}^{n} u_j^2 - 1\right) \qquad (8)$$

where λ is the Lagrange multiplier.

Differentiating (8) first with respect to x'_k results in the following expression for each k=1,...,n:

$$\sum_{\alpha=1}^{\gamma} ((\Delta_{\alpha k} - x'_k) - u_k \sum_{j=1}^{n} u_j (\Delta_{\alpha j} - x'_j)) = 0,$$

which can also be written as

$$\gamma M_k - \gamma x'_k = u_k \gamma t$$

where

$$M_k = \sum_{\alpha=1}^{\gamma} \Delta_{\alpha k} / \gamma \quad \text{and}$$

$$t = \sum_{\alpha=1}^{\gamma} (\sum_{j=1}^{n} u_j (\Delta_{\alpha j} - x'_j)) / \gamma.$$

Note that t does not depend on k. Thus for each k, $M_k = x'_k + tu_k$, which is just the statement that $\underset{\sim}{M}=(M_1,\ldots,M_n)$ lies on the line that minimizes $\sum_{\alpha=1}^{\gamma} p^2(\alpha)$. Hence the vector $\underset{\sim}{x}'$ can be taken as $\underset{\sim}{M}$.

Differentiating Equation 8 with respect to u_k, for k=1,...,n, yields

$$u_k\lambda - \sum_{\alpha=1}^{\gamma} (\sum_{j=1}^{n} u_j (\Delta_{\alpha j} - M_j)) (\Delta_{\alpha k} - M_k) = 0. \tag{9}$$

In terms of the statistics s_j, s_k, and r_{jk} defined earlier, this may be written

$$(\lambda / (\gamma - 1))u_k = \sum_{j=1}^{n} u_j s_j s_k r_{jk} , \tag{10}$$

for k = 1,...,n.

From (10) one may observe that $\underset{\sim}{u}$ is an eigenvector of the symmetric matrix

$$\begin{bmatrix} s_1^2 & s_2 s_1 r_{12} & \cdots & s_n s_1 r_{1n} \\ s_1 s_2 r_{12} & s_2^2 & \cdots & s_n s_2 r_{2n} \\ \vdots & \vdots & & \\ s_1 s_n r_{1n} & s_2 s_n r_{2n} & \cdots & s_n^2 \end{bmatrix} \qquad (11)$$

To determine which eigenvector, more information is needed about its associated eigenvalue, $\lambda/(\gamma - 1)$.

Multiplying each of the equations in (10) by its appropriate u_k and adding the resulting n equations together results in the following equation:

$$\lambda/(\gamma - 1) = \sum_{j=1}^{n} u_j^2 s_j^2 + 2 \sum_{1 \le j < k \le n} u_j u_k s_j s_k r_{jk} , \qquad (12)$$

using the fact that $\sum_{j=1}^{n} u_j^2 = 1$. The equation for $\sum_{\alpha=1}^{\gamma} p^2(\alpha)$ given in (7) can be written in terms of s_j, s_k, and r_{jk} as follows:

$$\sum_{\alpha=1}^{\gamma} p^2(\alpha) = (\gamma - 1) \sum_{j=1}^{n} s_j^2 - (\gamma - 1) \sum_{j=1}^{n} u_j^2 s_j^2 - 2(\gamma - 1) \sum_{1 \le j < k \le n} u_j u_k s_j s_k r_{jk}$$

Combining this expression with (12) yields

$$\sum_{\alpha=1}^{\gamma} p^2(\alpha) = (\gamma - 1) \sum_{j=1}^{n} s_j^2 - \lambda$$

To minimize the sum of the $p^2(\alpha)$ requires choosing the greatest value possible for the Lagrange multiplier λ. Thus, the desired eigenvector of the matrix (11) is that with maximum eigenvalue.

The $\underset{\sim}{x}'$ and $\underset{\sim}{u}$ that minimize $\sum_{\alpha=1}^{\gamma} p^2(\alpha)$ have now been expressed solely in terms of parameters derived from the set of $\underset{\sim}{\Delta}_{\alpha}, \alpha = 1,\ldots,\gamma$. The vector $\underset{\sim}{x}'$ is $\underset{\sim}{M}=(M_1,\ldots,M_n)$, and $\underset{\sim}{u}$ is the unit eigenvector of (11) with maximum eigenvalue. All the points of S were included in the calculation of L. When the line-fitting algorithm is illustrated, the line derived from all points is referred to as the preliminary line of best fit. As the algorithm is carried out, a number of intermediate lines may be calculated, each based on the points remaining after the removal of those whose distance from the previously calculated line exceeds some cutoff. The line ultimately resulting from application of the algorithm will be referred to as the line of best fit. When points are removed, the statistics given in (1), (2), and (3) must be recomputed, based on the remaining points. Lines calculated after the preliminary line are derived in the same way as the preliminary line, but with appropriately adjusted parameters.

Identifying and Analyzing Outlying Items

Let c be the distance from the line of best fit beyond which one considers a point's deviation possibly due to extraneous factors. This cutoff may be empirically determined, as is shown in the next section. The algorithm for calculating the line of best fit for a set $S=\{(\Delta_{\alpha 1},\ldots,\Delta_{\alpha n})\}_{\alpha=1,\ldots,\gamma}$ is as follows:

1. Find the line that best fits the set S-R, where initially R is the empty set.
2. Determine the distances of all points in S from the line. The distance of a given point, Δ_α, can be derived from (6), which can also be written as

$$p^2(\alpha) = \sum_{j=1}^{n} (\Delta_{\alpha j} - M_j)^2 - (\sum_{j=1}^{n} u_j (\Delta_{\alpha j} - M_j))^2 \quad (13)$$

3. Let R′ be the set of points whose distances from the line are greater than the cutoff distance c.
4. If R = R′, stop. Otherwise set R equal to R′ and begin again with step 1.

At the conclusion of the algorithm, the set R will contain those points considered outliers relative to the cutoff c. These items may be further analyzed by determining the groups contributing most to their deviance. For a given pair of groups, j and k, the line of best fit can be projected onto the j-k plane. For an outlying item α, the distance of $(\Delta_{\alpha j}, \Delta_{\alpha k})$ from the projected line can be calculated. A comparison of these distances over all possible pairs will reflect the relative contribution of the various pairs to the item's overall deviance.

The projection of the line of best fit onto the plane determined by the groups j and k is given by A+Bt, where $B = u_k/u_j$, $A = M_k - BM_j$, and t varies over the real numbers. The distance of $(\Delta_{\alpha j}, \Delta_{\alpha k})$ from the projected line can be derived from the two-dimensional analogue of (13). However, a more useful formula is provided by:

$$P_{jk}(\alpha) = (A - \Delta_{\alpha k} + B\Delta_{\alpha j}) / \sqrt{(1 + B^2)} \qquad (14)$$

This formula allows for positive and negative values and thus indicates whether $(\Delta_{\alpha j}, \Delta_{\alpha k})$ is above or below the projected line. The formula follows from the minimization of $(t-\Delta_{\alpha j})^2 + (A+Bt-\Delta_{\alpha k})^2$, which is just the square of the distance of $(\Delta_{\alpha j}, \Delta_{\alpha k})$ from a point $(t, A+Bt)$ on the projected line.

When (14) is positive it follows that $\Delta_{\alpha k} < A+B\Delta_{\alpha j}$. Thus $(\Delta_{\alpha j}, \Delta_{\alpha k})$ is below $(\Delta_{\alpha j}, A+B\Delta_{\alpha j})$, which means that it is below the projected line. In the same way it follows that when (14) is negative, $(\Delta_{\alpha j}, \Delta_{\alpha k})$ is above the projected line. A point below the line suggests that the item is unexpectedly difficult for group j. A point above the line suggests that the item is unexpectedly difficult for group k.

Illustration of Procedures

Item data for the examples presented below were selected from those used in Sinnott (1980). In this earlier work, the object of study was the Graduate Management Admission Test. A stratified sample of some 5,000 individuals taking the test in January 1977 provided the item data. Stratification was over a number of variables, including sex, age, ethnicity or race, and language fluency. Here groups varying in age are examined. Only one section of the test form is considered--problem solving. This was a 30-item, multiple-choice section presenting self-contained mathematical problems. An item's delta value is taken as its difficulty. This is just a linear transformation of the z score, given by $\Delta = 4z + 13$.

Figure 1 is a frequency distribution of the item distances from the preliminary line of best fit for a comparison involving three groups: randomly selected test takers between the ages 20 through 22, 35 through 39, and 40 through 65. There were about 1,450 individuals in the youngest group and 425 in each of the older groups. The numbers in Figure 1 refer to item numbers. Item 15 was the most deviant item, lying more than 1.6 delta units from the line.

Insert Figure 1 here

For the three-dimensional comparison, Figure 2 displays the results of calculating the line of best fit for a number of different item cutoffs. As smaller cutoffs are taken, the distribution of points within the cutoff distance begins to assume a configuration more consistent with the theoretically expected normal distribution of items about the line of best fit. The distribution presented at the top of the figure, is associated with the line in which items within 1.5 units are the only items participating in its derivation. For this cutoff, one iteration of the algorithm was required, since the line resulting from the removal of Item 15 was within 1.5 units of all the remaining points.

Insert Figure 2 here

In contrast, a cutoff of 1.0 resulted in the calculation of five lines, as illustrated in Table 1. The intercept and direction listed

21									
18									
13									
10					27	30			
8		17	28	26	23	25			
4	5	14	12	22	19	11		29	
3	1	4	6	16	9	2	20	24	15

0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0

Distance (delta units)

Figure 1. Frequency distribution of item distances from the preliminary line of best fit for the three-group comparison

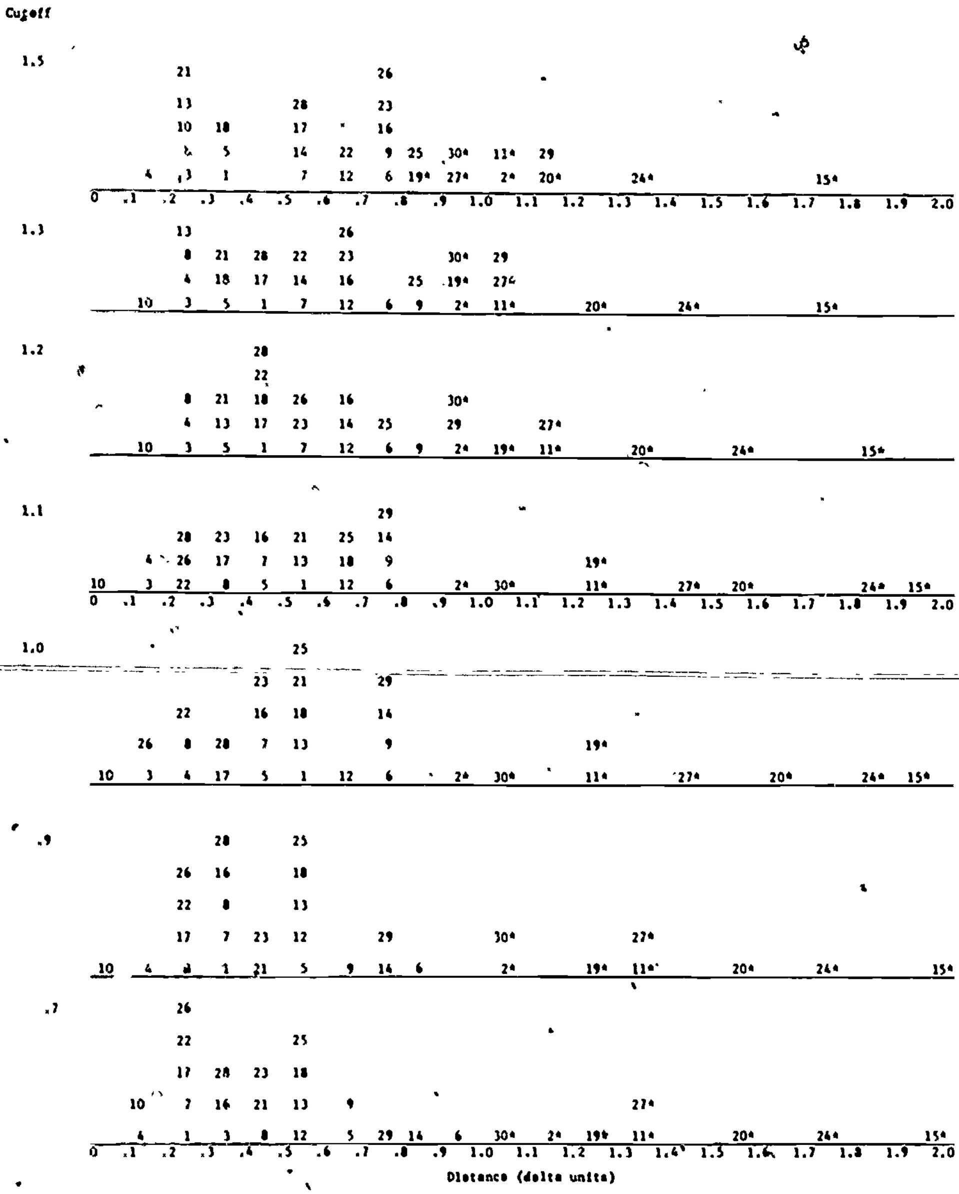

Figure 2. Frequency distribution of item distances as a function of cutoff for the three-group comparison

first are those associated with the preliminary line of best fit. Initially, items 15, 20, 24, and 29 were removed since they were greater than 1.0 units from the preliminary line. The reader may verify this by referring to Figure 1.

The line was then refitted relative to the remaining 26 points. The direction of the new line appears in Table 1 as (.60, .58, .54). The line contains the vector (13.65, 13.57, 13.72). All item distances were recalculated relative to the new line. In addition to the points previously removed, items 11 and 27 were greater than 1.0 units from the second line. Hence, for the next line derivation, items 11, 15, 20, 24, 27, and 29 were removed.

Item distances were calculated again. Of the removed points, only Item 29 was found to be within 1.0 units of the third line, and one additional point, Item 19, was more than 1.0 units from the line. Hence, the fourth line was based on the removal of items 11, 15, 19, 20, 24, and 27. In addition to these six items, Item 30 was more than 1.0 units from the fourth line. Hence, a fifth line was required. However, this was the final line, since those items greater than 1.0 units from it were the same as those previously removed. The distances given in Figure 2 for the distribution associated with the 1.0 cutoff are relative to this fifth and final line.

Insert Table 1 here

As can be seen in Figure 2, little new information about outliers is added as the cutoff drops below 1.1 delta units. By the 1.1 cutoff, the eight asterisked items have distinguished themselves relative to the main cluster of items, and these items also emerge as outliers when smaller cutoffs are taken.

TABLE 1

Results of the Algorithm
Applied to Three Groups

x'	u	Items removed
(13.92, 13.97, 14.08)	(.59, .59, .55)	15, 20, 24, 29
(13.65, 13.57, 13.72)	(.60, .58, .54)	11, 15, 20, 24, 27, 29
(13.53, 13.35, 13.50)	(.62, .58, .54)	11, 15, 19, 20, 24, 27
(13.63, 13.37, 13.49)	(.62, .57, .53)	11, 15, 19, 20, 24, 27, 30
(13.56, 13.32, 13.39)	(.63, .58, .53)	11, 15, 19, 20, 24, 27, 30

An examination of a plot of item difficulties for three groups randomly selected from the same pool of test takers, suggests that cutoffs below about 1.0 units are likely to remove items that deviate from the main cluster of points for reasons other than item inhomogeneity. Figure 3 displays the item distances from a preliminary line of best fit for a delta plot involving three groups of Caucasians, each with about 850 test takers. As can be seen, Item 29, the most deviant item, lies between .8 and .9 delta units from the line. Since the three groups were randomly sampled from the same population, this deviation cannot be attributed to factors that discriminate between the groups.

Insert Figure 3 here

For the age-group comparison, the line of best fit for the 1.0 cutoff was projected onto each of the two-dimensional planes defined by the different pairs of groups. For a given pair, j and k, the distance of $(\Delta_{\alpha j}, \Delta_{\alpha k})$ from the projected line was calculated for each item α in the set of outlying items. The results appear in Table 2. A positive value indicates that the item was unexpectedly difficult for the first group listed. A negative value indicates it was unexpectedly easy for the first group. As can be seen, all but one of the outlying items were unexpectedly difficult for older individuals when compared to younger test takers.

Insert Table 2 here

0–.1	.1–.2	.2–.3	.3–.4	.4–.5	.5–.6	.6–.7	.7–.8	.8–.9	.9–1.0
	28		30						
	19		27						
	18		14						
	11	23	13		24				
	10	17	12	26	21				
15	8	15	9	22	20	25			
4	2	1	7	5	3	6		29	

0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0

Distance (delta units)

Figure 3. Frequency distribution of item distances from the line of best fit for a comparison of three groups sampled from the same population

TABLE 2

Distance from Projected Lines
in a Three-Group Comparison

Item Number	Groups compared		
	35-39 vs. 20-22	40-65 vs. 20-22	40-65 vs.35-39
2	.8	.7	-.1
11	.7	1.3	.6
15	1.6	1.7	.2
19	1.0	1.1	.2
20	1.0	1.6	.6
24	1.7	1.4	-.2
27	1.3	1.1	-.1
30	- .4	.7	1.1

A fourth age group was added to the other age groups to yield a four-dimensional delta plot, the results of which are summarized in Figure 4. The fourth group was composed of about 575 test takers between the ages 30 and 34. The first distribution in Figure 4 displays the item distances from the preliminary line of best fit. With a few exceptions, there is considerable similarity in the distributions of Figures 2 and 4. Notable exceptions are items 25 and 30, both of which display considerably more deviant behavior in the four-dimensional analysis.

Insert Figure 4 here

For the four-dimensional analysis, Table 3 presents data similar to Table 2. Again, the line derived from the 1.0 cutoff was used. A comparison of Tables 2 and 3 reveals considerable overlap in their data. Among the new information emerging from Table 3 is the following. Item 30 was unexpectedly easy for the middle-age groups relative to both older and younger test takers, and Item 25 was unexpectedly easy for the 40- to 65-year-olds relative to the 30- to 34-year-olds.

Insert Table 3 here

A content analysis of outlying items may suggest reasons for their deviant behavior. However, information gathered from a single item must be interpreted cautiously since limitations in the methodology may lead to spurious data. Information extracted from a set of similarly behaving outliers is more reliable. However, there are shortcomings in this approach also, the foremost being that the reasons for an individual item's outlying behavior may be obscured in an aggregate analysis or

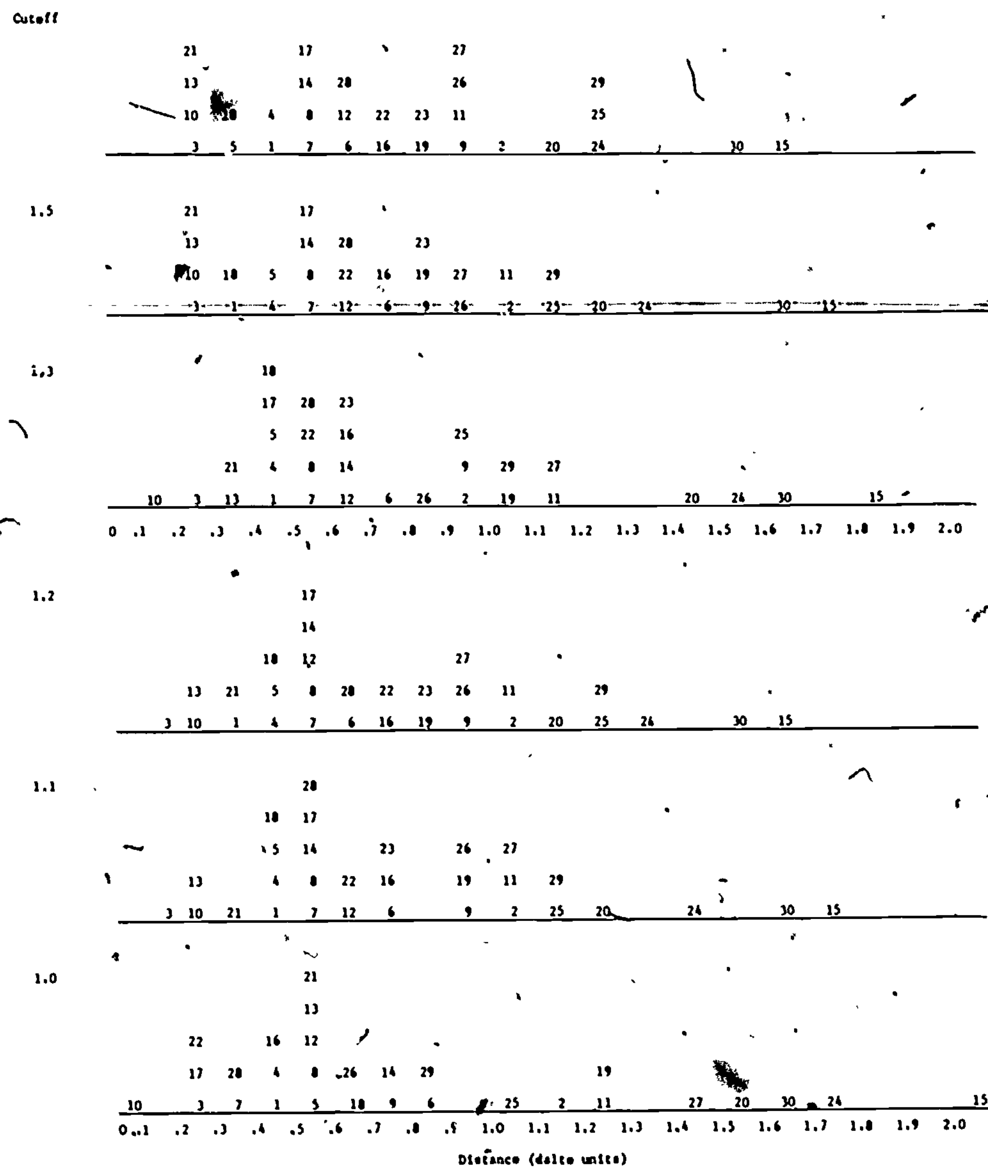

Figure 4. Frequency distribution of item distances as a function of cutoff for the four-group comparison

TABLE 3

Distance from the Projected Lines in a Four-Group Comparison

Item Number	Groups compared					
	30-34 vs. 20-22	35-39 vs. 20-22	40-65 vs. 20-22	35-39 vs. 30-34	40-65 vs. 30-34	40-65 vs. 35-39
2	.3	1.0	.9	.7	.6	0.
11	.7	.8	1.3	.1	.6	.5
15	1.0	1.7	1.8	.6	.8	.2
19	.6	1.0	1.1	.4	.5	.1
20	.6	1.0	1.5	.4	.9	.5
24	.9	1.7	1.3	.7	.4	- .2
25	.6	.1	- .5	-.5	-1.1	- .6
27	1.1	1.3	1.0	.1	0	- .2
30	-1.0	- .4	.7	.6	1.6	1.0

never pursued because other items sensitive to the same factors do not appear on the test.

In the age analysis, the set of items found to be unexpectedly difficult for older test takers did share a common characteristic. Of the thirty items in the problem-solving section, 17 were word problems, posing their questions in the context of some real-world situation. In contrast, the other 13 were more abstract, involving for the most part only mathematical concepts. Of the 13 non-word problems, eight appeared as unexpectedly difficult for 35- to 65-year-olds. The non-word problems deal with concepts seldom encountered in their purity outside of formal academic training. Their appearance as unexpectedly difficult for the older test takers may be due to a deterioration in a test taker's ability to manipulate these concepts, a deterioration correlated with the number of years elapsed since leaving school.

Discussion

In this paper a strategy has been presented for studying item bias using the intergroup comparison of item difficulties. In contrast to prior applications of this approach, the proposed method allows the simultaneous comparison of any number of groups, thus avoiding the awkwardness of numerous pair-wise comparisons. Fu .hermore, the role of deviant items in the formulation of criteria on which biased items are distinguished is lessened. The procedures outlined and illustrated in this paper allow for a more efficient and reliable application of item difficulty comparisons to the study of item bias.

The approach is based on an algorithm that ensures that the best-fitting line for an n-dimensional plot of item difficulties is derived solely from items lying within a specified distance of the line. The line that best fits these items is shown to be that which intersects the vector $\underset{\sim}{M} = (M_1, \ldots, M_n)$ and lies in the direction $\underset{\sim}{u}$, where $\underset{\sim}{u}$ is the unit eigenvector with maximum eigenvalue of the symmetric matrix given in (11) and M_j is the mean of the item difficulties experienced by group j.

It appears that no information is lost when additional dimensions are added to an analysis. Sinnott (1980) performed two-dimensional comparisons of the age groups studied here. The findings of that investigation were contained in the results of both the three- and four-dimensional comparisons. Furthermore, the findings of the three-dimensional analysis are repeated in the four-dimensional analysis, as can be seen by comparing Tables 2 and 3.

The limitations in the proposed strategy stem primarily from the assumption underlying the use of the inverse normal transformation as a measure of item difficulty. One bases the use of this transformation on the existence of a level of ability above which success on the item is achieved and below which, failure. No item in practice has such perfect discrimination. A more accurate model assumes that test takers over the range of ability may respond correctly to the item, but their chances of correct response improve with ability. Using this model, Lord (1977) has illustrated how items sensitive to the same dimension may deviate from linearity in a plot of their z scores.

Empirically, the algorithm seems to be a useful tool for the study of item bias. Several theoretical issues remain to be explored, however. In particular, the resulting line may be only one of several lines that satisfy the criterion of resulting from the consideration of all points lying within a specified distance. One might wish to show that the algorithm uncovers that line with the most points contributing to its calculation. Also, it is theoretically possible that the algorithm might never lead to a solution, but circle endlessly. However, this seems unlikely to happen on any real data set.

References

Angoff, W. H., & Ford, S. F. Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 1973, 10, 95-105.

Donlon, T. F., Hicks, M. M., & Wallmark, M. M. Sex differences in the item responses on the Graduate Record Examination. Applied Psychological Measurement, 1980, 4, 9-20.

Lord, F. M. A study of item bias, using item characteristic curve theory. In Y. H. Poortinga (Ed.), Basic problems in cross-cultural psychology. Amsterdam and Lisse: Swets and Zeitlinger B. V., 1977.

Pearson, K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 1901, 2, 559-572.

Rudner, L. M., Getson, P. R., & Knight, D. L. Biased item detection techniques. Journal of Educational Statistics, 1980, 2, 213-233.

Sinnott, L. T. Differences in item performance across groups. Research Report 80-19. Princeton, NJ: Educational Testing Service, 1980.