DOCUMENT RESUME

ED 275 732 TM 860 646

AUTHOR	Reckase, Mark D.; And Others
TITLE	When Unidimensional Data Are Not Unidimensional.
INSTITUTION	Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE	Jun 85
CONTRACT	N00014-81-K0817
NOTE	24p.; Paper presented at the Annual Meeting of the Psychometric Society (50th, Nashville, TN, June 1-4, 1985). Some figures contain small, broken type.
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Correlation; Difficulty Level; *Factor Analysis; *Item Analysis; *Latent Trait Theory; Mathematical Models; *Statistical Distributions; Statistical Studies; Testing Problems; Test Items; True Scores
IDENTIFIERS	*Dichotomous Variables; *Tetrachoric Correlation

ABSTRACT

Factor analysis is the traditional method for studying the dimensionality of test data. However, under common conditions, the factor analysis of tetrachoric correlations does not recover the underlying structure of dichotomous data. The purpose of this paper is to demonstrate that the factor analyses of tetrachoric correlations is unlikely to yield clear support for unidimensionality even when the data are generated to be unidimensional. This result is caused by a failure of the item data to meet the assumptions of the tetrachoric correlation. For this study, item true score distributions were generated assuming a normal latent trait and a variety item characteristic curve (ICC) forms for the items. In every case, these distributions were nonnormal, and the bivariate distribution did not match the bivariate normal. The principal component analysis of data generated according to these ICC's yielded a highly complex solution, most likely a result of the violation of the assumptions of the tetrachoric correlations that form the basis of the analysis. Further research is needed on new methods of factor analysis of dichotomous test data generated by a variety of ICC forms. (BS)

When Unidimensional Data Are Not Unidimensional

Mark D. Reckase

James E. Carlson

Terry A. Ackerman

The American College Testing Program

Factor analysis has been the traditional method for studying the dimensionality of test data. This is true for dichotomous data even though several authors have documented problems with the application of factor analysis to this type of data (Dingman, 1958; Ferguson, 1941; Gourlay, 1951; Guilford, 1941; McDonald and Ahlawat, 1974). The continued use of factor analysis, especially with tetrachoric correlations, for the analysis of dichotomous data probably stems from the need to verify the unidimensionality assumption required for many item response theory (IRT) models. In addition, Lord and Novick (1968) suggest that the analysis of tetrachoric correlations may be helpful in supporting the assumption, even though they exhibit appropriate caution in their discussion of the topic.

However, under fairly common conditions, the factor analysis of tetrachoric correlations does not recover the underlying structure of dichotomous data (Gourlay, 1951; Reckase, 1979). This paper presents some

---

Paper presented at the meeting of the Psychometric Society, Nashville, TN, June 1985. This research was supported by Contract Number N00014-81-K0817 from the Personnel and Training Research Programs of the Office of Naval Research.

reasons why this should be the case if it can be assumed that the dichotomous data can be accurately described by an IRT model. Specifically, this paper will show that the assumptions of the tetrachoric correlation are not consistent with a general class of IRT models. The relationship between the IRT models for two test items and the bivariate distribution of the ability to respond to two test items will be described first. This relationship will then be used to discuss the tetrachoric correlations between two items and the implications these correlations have for factor analyses of dichotomous test data.

A Model of the Relationship between Scores on Dichotomous Items and a Hypothetical Latent Trait

In this paper it is assumed that the relationship between the performance of a person on a test item and the trait measured by the item is so complex that it can only be described by a probabilistic model. The probabilistic model is defined by a function that relates the probability of a correct response to the item to the level of ability of a person on a hypothetical latent trait. This function may be described either by a mathematical formula or by a set of ordered pairs of probabilities and corresponding abilities. For this paper, the probabilistic model will be specified by the set of ordered pairs because it defines a more general class of IRT models than can be defined by mathematical formulas.

According to this model, for each value of the latent ability being measured by an item, there is a corresponding probability of a correct response to the item. The fact that a probabilistic model is being used implies that there is uncertainty about the response of the person to the item. At different times and under different conditions, different responses may be given to the same item by the same person.

One way to explain the probabilistic relationship between latent ability and the item score is to assume that the ability to respond correctly to an item is a function of a very large number of variables that describe the mental state of the person taking the item. Since each state variable accounts for a very small proportion of the variance of the item score, and because there are very many variables, the result can only be described by a distribution of uncertainty for the individual on the item trait. Lord and Novick (1968) have called this distribution a propensity distribution. Thurstone (1927) called it a discriminal dispersion. Because the distribution is based on the effects of the combination of a large number of variables, it can be assumed to be normal.

The propensity distribution is defined on the scale of the ability that is required to respond correctly to the item. Whether or not a person obtains a correct response to the item depends on whether or not their ability is above or below a critical value for the item. The critical value is located at a point that divides the distribution into two parts, the upper part containing a proportion equal to the probability of a correct response and the lower part corresponding to the probability of an incorrect response.

The mean of the propensity distribution for a person's response to an item can be determined from the person's ability and the IRT function. Using the ability and the IRT function, the probability of a correct response can be determined. The inverse normal distribution function can then be applied to the probability to obtain the corresponding z-score. If the critical value of the item is arbitrarily set at zero (this can be done because the origin of the scale is undetermined), the z-score is equal to the mean of the propensity distribution for that person on that item. Since the mean of the propensity distribution has been defined as the true score by Lord and Novick (1968), this process also defines the true score for a person on an item. The process of conversion, from latent trait to true score on the item scale, is summarized in Figure 1.

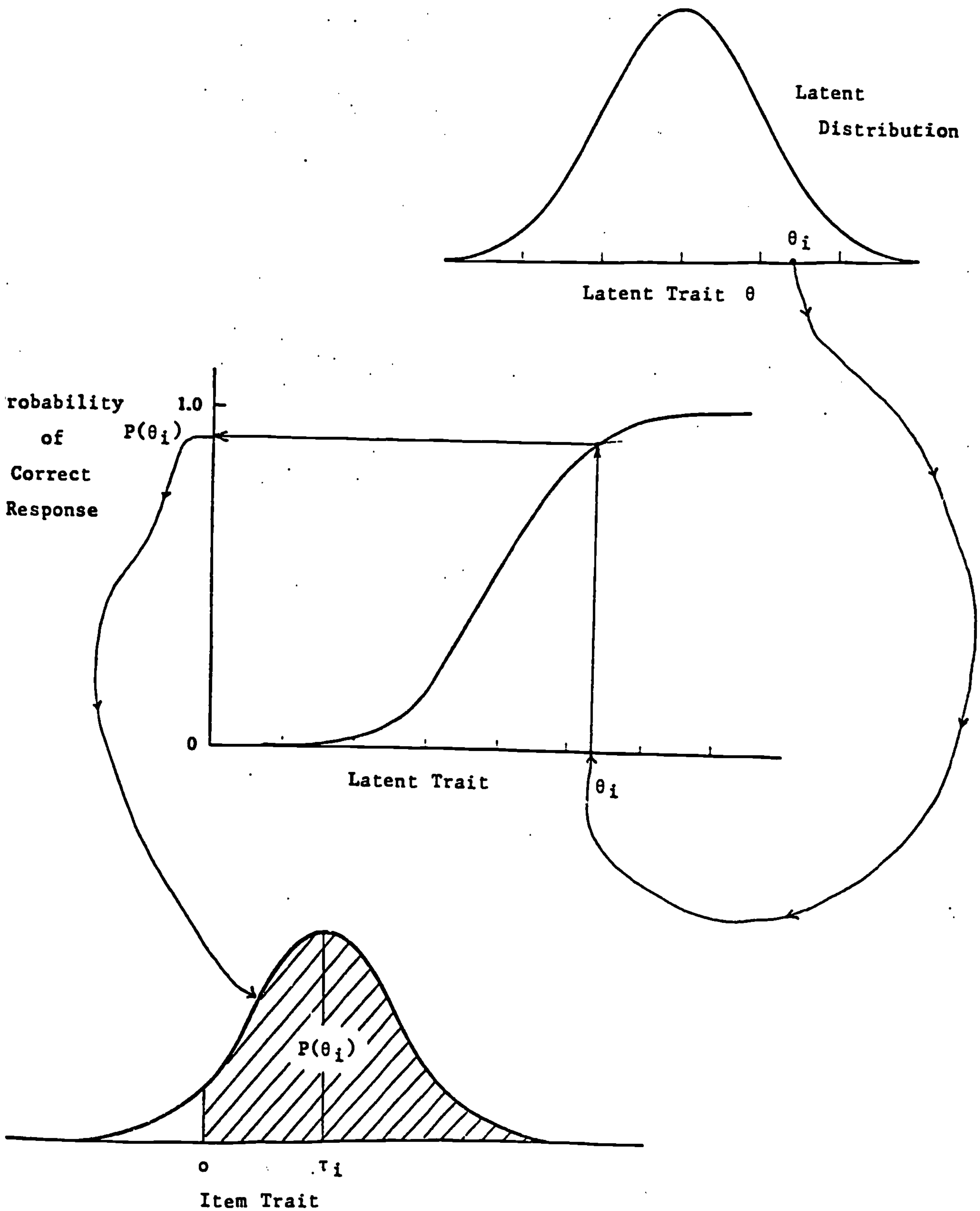

Figure 1

Conversion from Latent Trait to True Score on the Item Scale

By transforming all of the abilities in the latent distribution to means of propensity distributions, the distribution of true scores on the item trait can be determined. If this is done for two items simultaneously, the bivariate distribution of the true scores on the items can be determined.

Item Trait Distributions Implied by Several ICC Models

In order to determine the characteristics of the distribution of true scores on the item traits given that the distribution on the latent trait is standard normal (N (0, 1)), 2,000 cases were generated using the IMSL (1980) random normal number generator. For each of these values, the probability of a correct response to a series of hypothetical items was determined from the ICC's for the items. The ICC's for the items were specified by ordered pairs of the probabilities that corresponded to θ-values of -3, -2, -1, 0, 1, 2, 3. The probability of a correct response for the 2,000 cases was determined by linear interpolation or extrapolation if the values did not correspond to the seven values used to specify the probabilities. Once the probabilities were determined, the true scores on the item scales were obtained using the inverse normal transformation.

The distributions of item traits were obtained for three different ICC models. The probabilities corresponding to the seven θ-values for the three items are given in Table 1.

Table 1

Probabilities Defining the ICC's for Three Items

Item	θ-Value						
	-3	-2	-1	0	1	2	3
2	.10	.05	.20	.55	.70	.80	.90
17	.15	.15	.15	.15	.30	.40	.60
20	.50	.40	.20	.50	.70	.80	.90

Item 2 is a moderately difficult item with a lower asymptote of .10. This item has a slightly nonmonotonic item characteristic curve (ICC). The item true score distribution that corresponds to the latent distribution for this item is given in Figure 2a. As can be seen, this distribution is negatively skewed with a skewness of -.58. The item true score distribution that corresponds to the latent trait distribution for Item 17, a very hard item, is given in Figure 2b. This distribution is highly positively skewed (skewness = 1.32). Item 20 is a moderately difficult item with a strongly nonmonotonic item characteristic curve. The item true score distribution for this item is shown in Figure 2c. This distribution also deviates substantially from a normal distribution. However, in this case the deviation is in the form of being platykurtic (kurtosis = -.865).

ITEM2

```
    COUNT   MIDPOINT   ONE SYMBOL EQUALS APPROXIMATELY 10.00 OCCURRENCES

        0     -2.0
        0     -1.8
       37     -1.6   :***
       45     -1.4   **:**
       53     -1.2   ****:
      124     -1.0   *******:****
      126      -.8   ************:
      113      -.6   ***********        .
      159      -.4   ****************     .
      154      -.2   ***************        .
      172       .0   *****************       .
      407       .2   ***********************:*****************
      273       .4   *****************:********
      216       .6   ************:********
       93       .8   ********:
       23      1.0   **    .
        2      1.2     .
        1      1.4   .
        2      1.6   .
        0      1.8
        0      2.0
                     I....+....I....+....I....+....I....+....I....+....I
                     0        100       200       300       400       500
                               HISTOGRAM FREQUENCY
```

MEAN	-.075	STD ERR	.014	MEDIAN	.118
MODE	.260	STD DEV	.610	VARIANCE	.373
KURTOSIS	-.439	S E KURT	1.999	SKEWNESS	-.501
S E SKEW	.055	RANGE	3.216	MINIMUM	-1.637
MAXIMUM	1.579	SUM	-150.228		

Figure 2a

Item True Score Distribution for Item 2

ITEM17

```
COUNT   MIDPOINT   ONE SYMBOL EQUALS APPROXIMATELY 24.00 OCCURRENCES

 1043    -1.050    *****:*************************************
  122     -.975    *****     .
   96     -.900    ****       .
  111     -.825    *****      .
  105     -.750    ****      .
   93     -.675    ****     .
   82     -.600    ***    .
   86     -.525    ***:
   96     -.450    **:*
   70     -.375    *:*
   44     -.300    :*
   19     -.225    :
   12     -.150    *
   14     -.075    *
    1      .000
    1      .075
    0      .150
    2      .225
    1      .300
    0      .375
    2      .450
                   I....+....I....+....I....+....I....+....I....+....I
                   0        240       480       720       960      1200
                             HISTOGRAM FREQUENCY
```

MEAN	-.851	STD ERR	.006	MEDIAN	-1.036
MODE	-1.036	STD DEV	.256	VARIANCE	.066
KURTOSIS	1.111	S E KURT	1.999	SKEWNESS	1.321
S E SKEW	.055	RANGE	1.520	MINIMUM	-1.036
MAXIMUM	.484	SUM	-1701.402		

Figure 2b

Item True Score Distribution for Item 17

ITEM20

```
   COUNT   MIDPOINT   ONE SYMBOL EQUALS APPROXIMATELY  8.00 OCCURRENCES

       0     -1.15   .
       0     -1.00    .
      87      -.85  ******:****
     204      -.70  ***********:*************
     174      -.55  *****************:****
     174      -.40  *********************.
     167      -.25  *********************
     168      -.10  *********************
     252       .05  ******************************:*
     221       .20  ***************************:
     163       .35  ********************
     143       .50  ****************:*
     143       .65  **********:*******
      67       .80  ******:*
      31       .95  ***:
       1      1.10   .
       2      1.25  .
       1      1.40
       2      1.55
       0      1.70
       0      1.85
                    I....+....I....+....I....+....I....+....I....+....I
                    0        80       160       240       320       400
                              HISTOGRAM FREQUENCY

MEAN          -.035     STD ERR       .011     MEDIAN       -.006
MODE           .177     STD DEV       .476     VARIANCE      .227
KURTOSIS      -.865     S E KURT     1.999     SKEWNESS      .131
S E SKEW       .055     RANGE        2.420     MINIMUM      -.841
MAXIMUM       1.579     SUM        -70.148
```

Figure 2c

Item True Score Distribution for Item 20

The bivariate true score distributions for each of the pairs of items are given in Figures 3a, 3b, and 3c. For all of the cases shown here, the bivariate distribution of the item traits is a tight curve. Clearly the assumption of linearity is not supported. However, the strength of the relationship clearly demonstrates the unidimensional nature of these data.

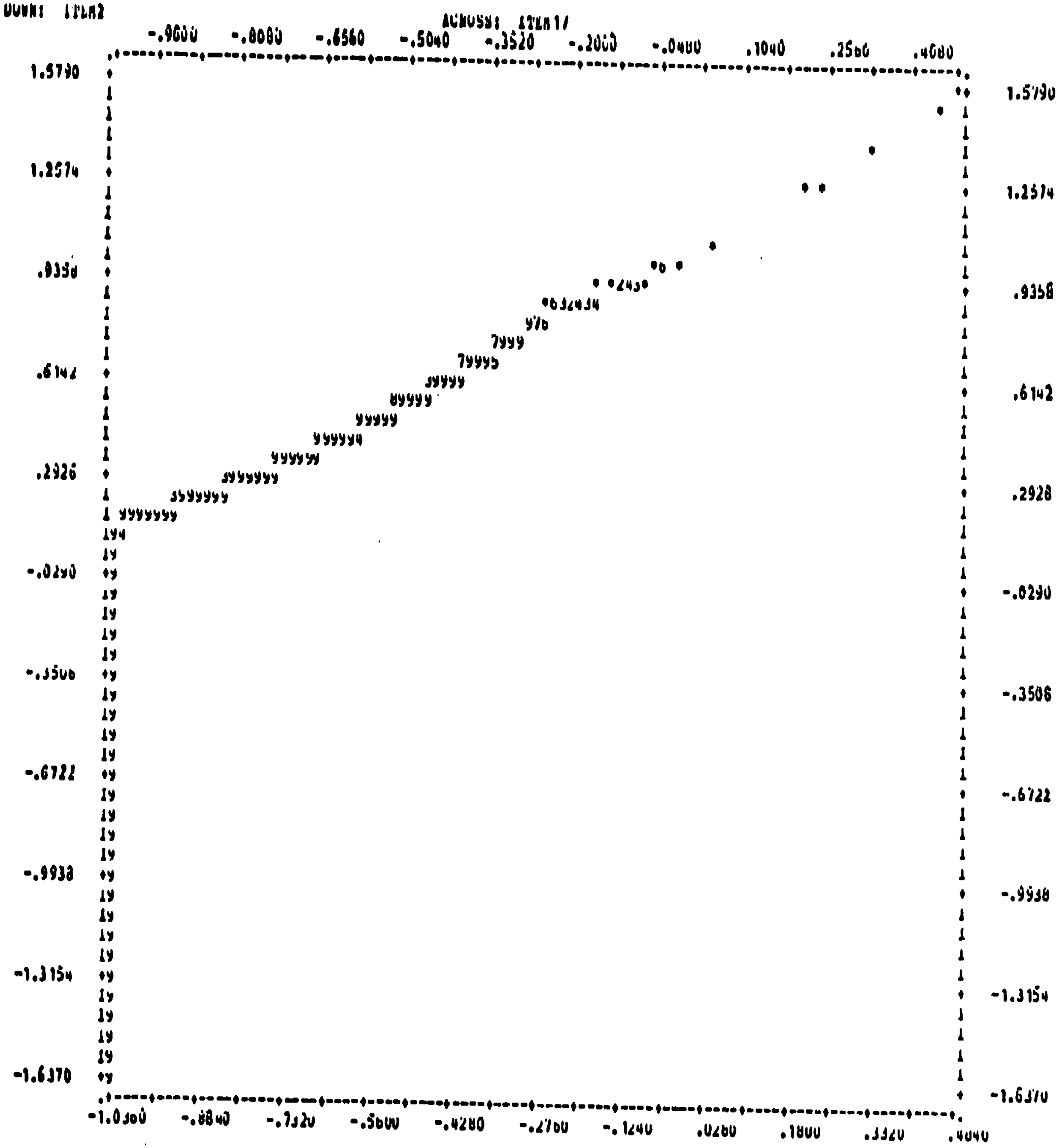

Figure 3a

Bivariate Distribution for Item True Scores

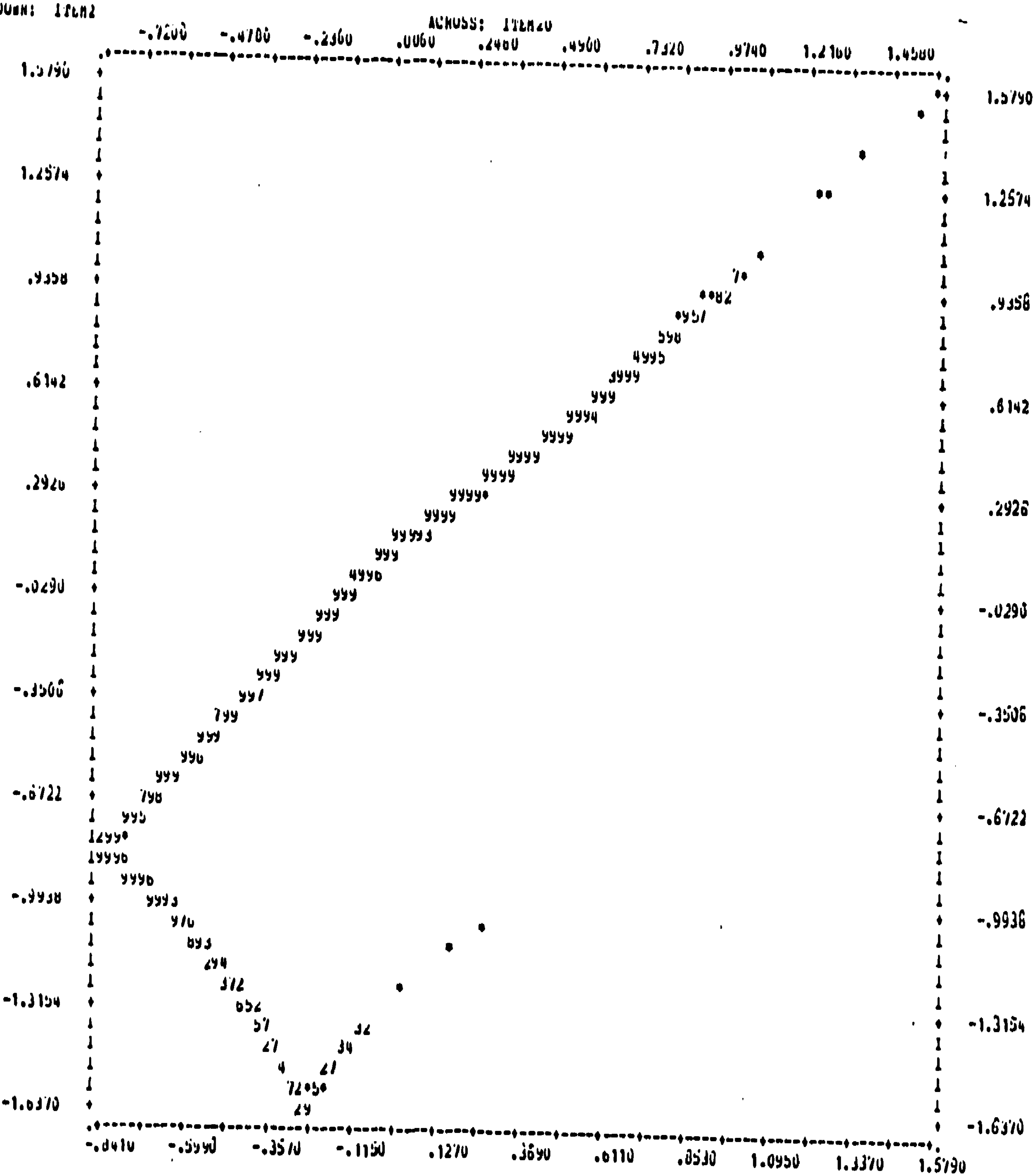

Figure 3b

Bivariate Distribution for Item True Scores

However, true scores are never observed. To obtain the continuous score equivalent of the observed item scores, scores were randomly sampled from the propensity distributions for each person-item combination. The bivariate observed score distributions for the three items given in Table 1 are presented in Figures 4a, 4b and 4c. These are the distributions whose ρ-parameter is estimated by the tetrachoric correlation coefficient. Note that these distributions are not bivariate normal.

Propensity Distribution

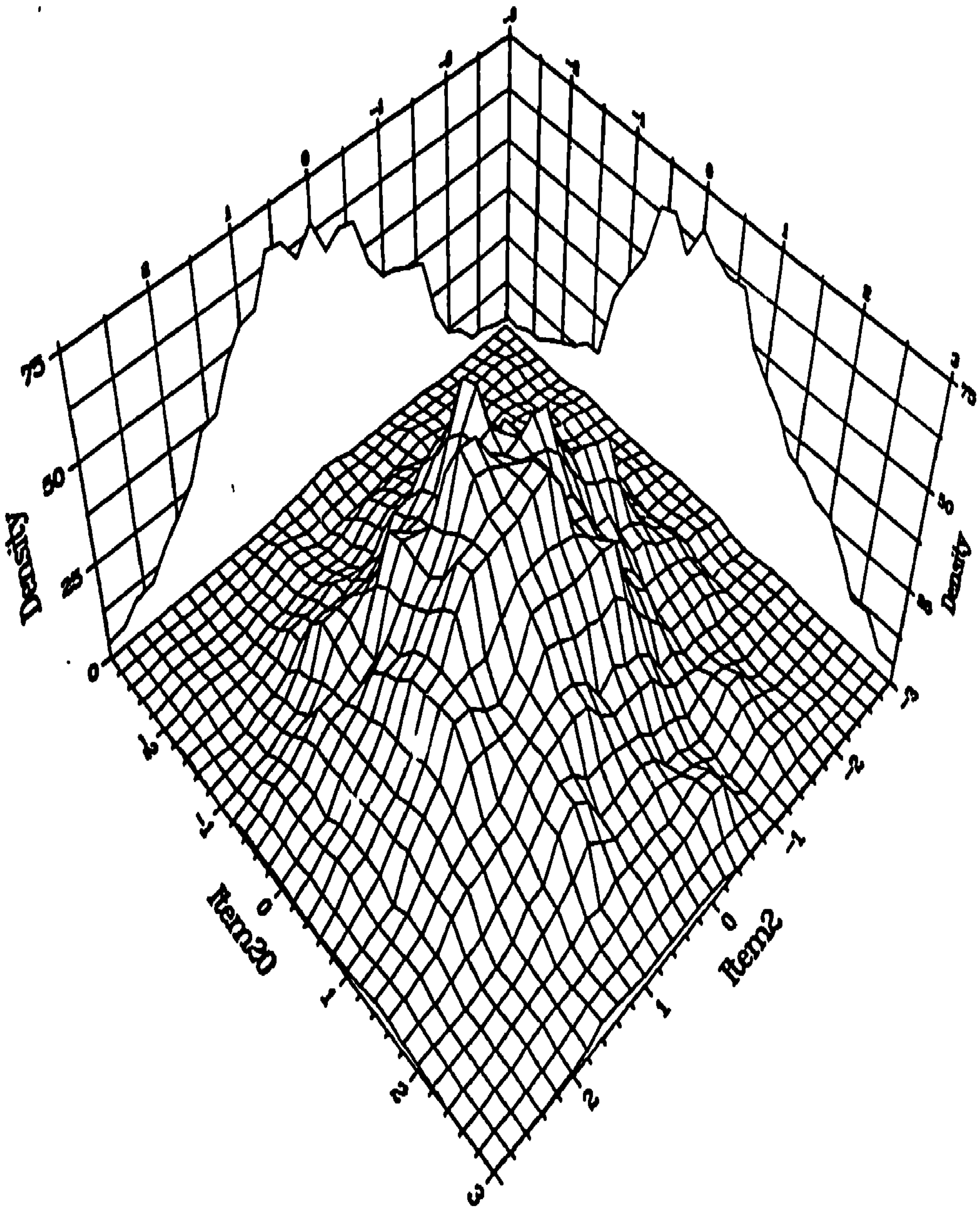

Figure 4a

Bivariate Item Score Distribution

Propensity Distribution

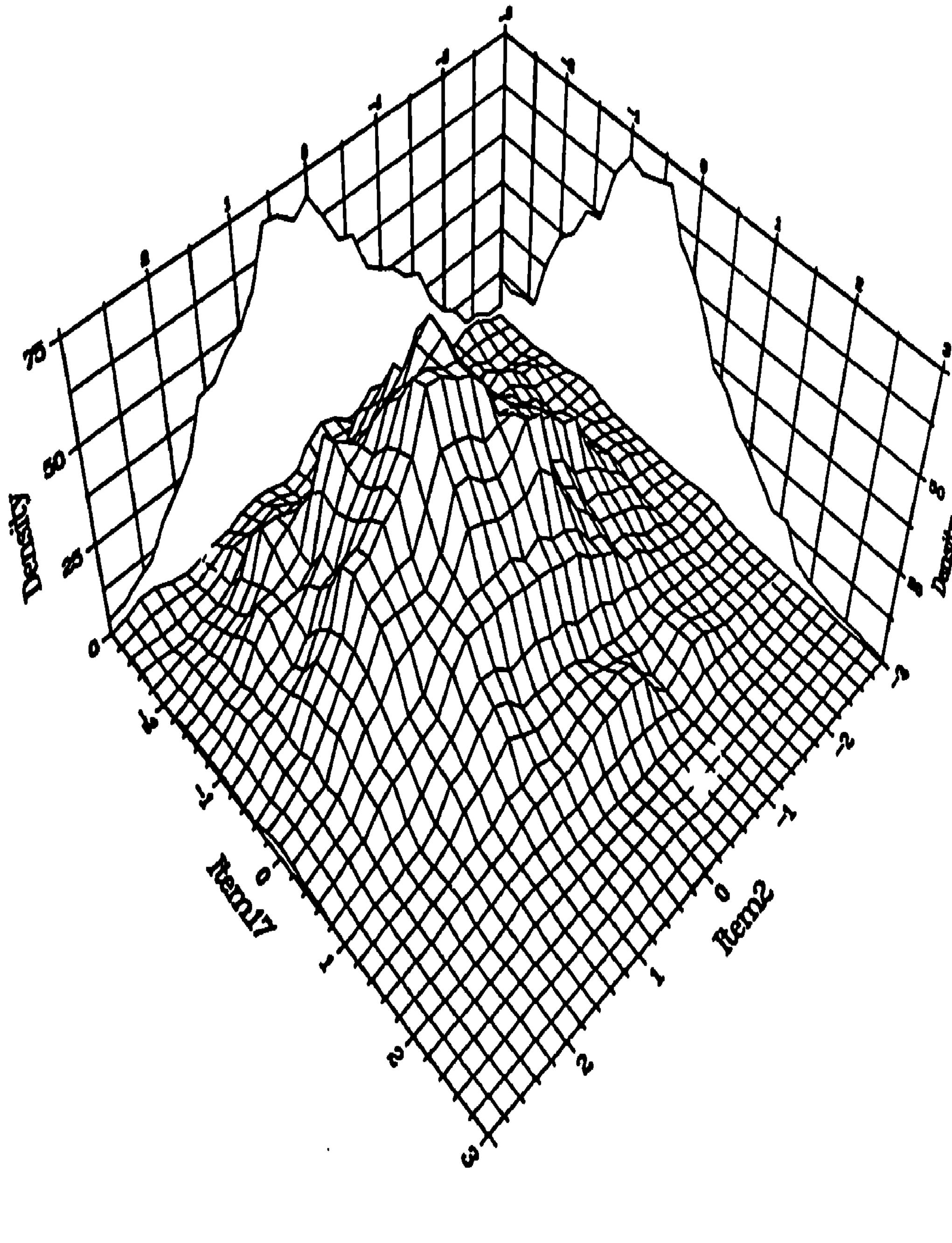

Figure 4b

Bivariate Item Score Distribution

Propensity Distribution

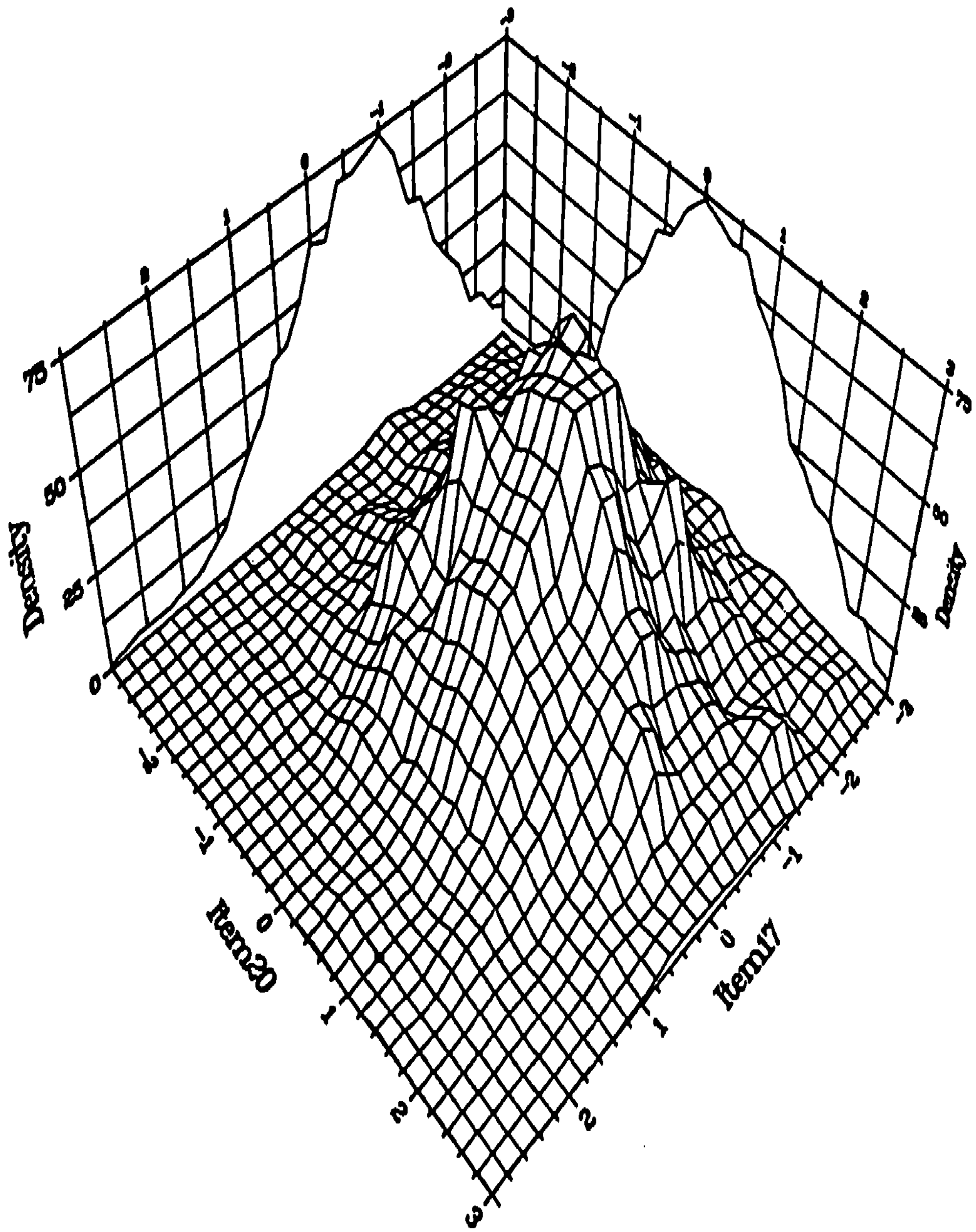

Figure 4c

Bivariate Item Score Distribution

Factor Analysis Results

In order to demonstrate the effects of violating the assumptions of the tetrachoric correlations on factor analyses, dichotomous data were generated using many different types of ICC's. The probabilities used to describe these ICC's are given in Table 2. The factor loading matrix and eigenvalues from the principal component analysis of the tetrachoric correlations for these data are given in Table 3.

Table 2

Probabilities Corresponding to Seven Ability Levels for Twenty Hypothetical Items

Item	Ability Level						
	-3	-2	-1	0	1	2	3
1	00	80	85	90	92	95	98
2	10	05	20	55	70	80	90
3	10	30	70	80	90	95	99
4	10	10	40	70	80	90	95
5	10	10	15	50	70	80	90
6	50	70	90	91	92	93	97
7	40	60	80	90	95	97	99
8	35	50	70	90	95	97	99
9	20	40	60	80	90	95	99
10	15	20	50	70	90	95	99
11	15	15	40	60	80	90	95
12	20	15	30	50	70	90	95
13	15	15	20	40	60	80	90
14	20	15	15	30	50	70	90
15	15	15	15	15	40	60	90
16	20	15	15	15	40	50	80
17	15	15	15	15	30	40	60
18	25	20	15	15	15	30	50
19	00	00	40	40	60	60	90
20	50	40	20	50	70	80	90

Note: Decimal points have not been included. All values are to two decimal places.

Table 3

Unrotated Principal Components of the Tetrachoric Correlations

Item	Component Loadings*				
	1	2	3	4	5
1	.20		.28	.36	.59
2	.28	.52		.25	-.41
3	.51	-.59			
4	.49	.56	-.24		
5	.58		.48		
6	.67		-.22		
7	.38	.66			
8	.27		.76	-.24	
9	.49			.52	-.22
10	.56				
11	.58			.31	
12	.63				
13	.61				
14	.59				
15	.56				
16	.52		-.22	-.25	
17	.49			-.30	.39
18	.41			-.41	-.28
19	.34		-.36	-.29	
20	.24	.44	.37	-.27	.31
Eigen value	4.76	1.71	1.47	1.18	1.02

Note: *Loadings less than .2 in absolute value have been deleted.

As can be seen from this analysis, the unidimensional nature of the ability dimension was not supported. Five factors are present with eigenvalues greater than 1.0 and none of the factors are readily related to item characteristics.

Discussion and Conclusions

The purpose of this paper was to demonstrate that the factor analysis of tetrachoric correlations is unlikely to yield clear support for unidimensionality even when the data are generated to be unidimensional. This result is caused by a failure of item data to meet the assumptions of the tetrachoric correlation.

In this study, item true score distributions were generated assuming a normal latent trait and a variety of forms for the ICC's for the items. In every case, these distributions were shown to be nonnormal, and the bivariate distributions were shown not to match the bivariate normal. The principal component analysis of data generated according to these ICC's yielded a highly complex solution, most likely a result of the violation of the assumptions of the tetrachoric correlations that form the basis of the analysis.

New methods for factor analysis have recently been developed specifically for dichotomous data (Bock and Aitken, 1981; McDonald, 1967; Muthén, 1983; Christoffersson, 1981). These methods may be better able to meet the requirements of data of this type. However, these methods assume a particular form for an ICC and they may not be able to accurately describe data that are generated using a different form for an ICC. This is clearly an area for future research.

References

Bock, R. D. and Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. Psychometrika, 46, 443-459.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. Psychometrika, 40, 5-32.

References - Cont.

Dingman, H. F. (1958). The relation between coefficients of correlation and difficulty factors. British Journal of Statistical Psychology, 11, 13-17.

Ferguson, G. A. (1941). The factorial interpretation of test difficulty. Psychometrika, 6, 323-329.

Gourlay, N. (1951). Difficulty factors arising from the use of tetrachoric correlations in factor analysis. British Journal of Psychology (Statistical Section), 65-76.

Guilford, J. P. (1941). The difficulty of a test and its factor composition, Psychometrika, 6, 67-77.

IMSL (1980). IMSL library reference manual, edition 8. Houston: IMSL Inc.

Lord, F. M. and Novick, M. R. (1968). Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley.

McDonald, R. P. (1967). Nonlinear factor analysis. Psychometric Monographs, 15.

McDonald, R. P. and Ahlawat, K. S. (1974). Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 34, 273-286.

References - Cont.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables Psychometrika, 43, 551-560.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor data: results and implications. Journal of Educational Statistics, 4(3), 207-230.

Thurstone, L. L. (1927). A law of comparative judgement. Psychological Review, 34. 273-286.