DOCUMENT RESUME

ED 402 315 TM 025 806

AUTHOR	Dirir, Mohamed A.; Sinclair, Norma
TITLE	On Reporting IRT Ability Scores When the Test Is Not Unidimensional.
PUB DATE	Apr 96
NOTE	26p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Ability; *Error of Measurement; *Estimation (Mathematics); *Item Response Theory; *Scores; Simulation
IDENTIFIERS	*Dimensionality (Tests); Stouts Procedure; *Unidimensionality (Tests)

ABSTRACT

The purpose of this study was to examine the effect of test dimensionality on the stability of examinee ability estimates and item response theory (IRT) based score reports. A simulation procedure based on W. F. Stout's Essential Unidimensionality was used to generate test data with one dominant trait for the whole test and three minor traits specific to three subsets of items. The dimensionality of the data was controlled by varying the relative strengths of the specific traits. The errors in the ability estimation, which were examined both at test level and at subtest level, were compared among different degrees of test dimensionality. The correlation between the dominant trait and the minor traits was varied to three levels. When major and minor traits were not correlated, the standard errors in the ability estimates increased with increase in the strength of the minor traits. When the major and minor traits were correlated, on the other hand, the errors in the ability estimates slightly decreased as the strength of the minor traits was increased. (Contains 2 figures, 5 tables, and 12 references.) (Author/SLD)

On Reporting IRT Ability Scores When the Test is not Unidimensional

Mohamed A. Dirir

Norma Sinclair

Connecticut State Department of Education

A paper presented at the annual meeting of the National Council on Measurement in Education

New York

April 1996

Abstract

The purpose of this study was to examine the effect of test dimensionality on the stability of examinee ability estimates and IRT-based score reports. A procedure based on Stout's Essential Unidimensionality was used to generate test data with one dominant trait for the whole test and three minor traits specific to three subsets of items. The dimensionality of the data was controlled by varying the relative strengths of the specific traits. The errors in the ability estimation, which were examined both at test level and at subtest level, were compared among different degrees of test dimensionality. The correlation between the dominant trait and the minor traits was varied to three levels.

When major and minor traits were not correlated, the standard errors in the ability estimates increased with increase in the strength of the minor traits. When the major and minor traits were correlated, on the other hand, the errors in the ability estimates slightly decreased as the strength of the minor traits was increased.

On Reporting IRT Ability Scores

When the Test is not Unidimensional

Several testing institutions use item response theory (IRT) in test development, test scoring, and test score reporting. In scoring a test an examinee's ability is sometimes estimated using known item parameters. In such cases predictions can be made about the performance of an examinee with a certain ability on any set of items in a precalibrated item pool. Item parameters are often unknown in typical testing situations, and examinee abilities and item parameters are jointly estimated. Examinee scores are then reported on the ability scale as estimated in the joint calibration. Alternately, linear or nonlinear transformation of the ability scores, which are more convenient to interpret scores, may be reported (Hambleton, Swaminathan, & Rogers, 1991).

Regardless of the scale on which scores are reported, test developers either report scores on the whole test or report scores in subtests or clusters depending on the intended uses of the assessment. The standard errors of estimates for each ability score may also be added to the score reports, as recommended in the Standards for Educational and Psychological Testing (1985), to further enhance the accuracy of score interpretations.

Examples of situations where transformed IRT ability scores are reported are NAEP and the 10th grade Connecticut Academic

Performance Assessment. Among the advantages of using IRT in test reporting are test score equating and administration of computer adaptive tests. However, these advantages are fully achieved only when several assumptions of the theory are met and the model of choice fits the test data.

One assumption, which is critical but not often fully met, is the assumption of unidimensionality which requires that test items measure a single underlying construct. Practically speaking, although the test might have been initially designed to measure a single trait, a given set of test items will almost certainly not be strictly unidimensional (see, for example, Traub, 1983). In the case of mastery tests, for example, a test may not be unidimensional because it is designed to measure different objectives or clusters in a single subject area. The items for each objective or cluster may in turn be influenced by a trait specific to that objective. That could also be true for licensure or credentialing examinations. According to Shea, Norcini and Webster (1988, p.285) "Licensure and certification tests in the professions may pose a special challenge to the implementation of IRT methods...expertise may be required in many areas so as to seriously challenge the IRT assumption of uni-dimensionality".

While it is true that the assumption of unidimensionality may be difficult to meet, several researchers have noted that tests often have a dominant dimension and several minor dimensions (See, for example, Drasgow & Parsons, 1983; Harrison, 1986; Stout, 1987; Nandakumar, 1991), and when the minor dimensions are relatively unimportant, the test may be assumed to be unidimensional. It has been established that unidimensional IRT models and programs could be practically used in such situations. In light of this, Stout (1987) has proposed the concept of essential unidimensionality which holds when these minor traits are less potent (Stout,1987; Nandakumar, 1991). As the minor traits become stronger, the test departs from being essentially unidimensional. This violation of the unidimensionality assumption may be expected to adversely affect the reliability and validity of the test. If this is the case, it will have important implications for testing practitioners. One possible outcome is large standard errors in the examinee ability estimates, and hence inaccuracy in the examinee score reports.

The purpose of this study was to investigate how various degrees of test dimensionality impact on examinee ability estimates and score reports. Score reporting at subtest level and reporting at test level were examined in light of the standard errors of ability scores at certain degrees of test dimensionality. By using simulated data of known dimensionality, and then proceeding as if the assumption of unidimensionality were met, it was possible to evaluate the effect of violating the unidimensionality assumption on the ability estimates and test score reports.

Method

Test data with different degrees of dimensionality were simulated. The dimensionality structure was modeled using the procedure reported in Stout (1987) and Nandakumar (1991). In each simulated data set, there was one major trait and three minor traits. The major trait was set to influence all test items while each of the minor traits was set to influence a subset of items. Also, the minor traits influenced equal numbers of items, and their potencies were set to be equal. The dominant trait was intended to reflect broad subject area proficiency such as mathematics. On the other hand, the minor traits were intended to represent specific domains within a subject area such as algebra, geometry, and probability in a mathematics test. In practical testing situations one would anticipate some relationship among the general and specific traits. Hence, three levels of correlation between the dominant trait and the minor traits were examined.

The amount of dimensionality in the data sets were controlled by the strength of the minor traits relative to the major trait using the procedure proposed by Stout (1987). In Stout's procedure, an index ξ controls the mean and variance of the discrimination parameter along the major and minor traits, and hence the relative strengths of the traits. The relationship between the discrimination parameters can be written as (Nandakumar, 1991):

$$a_1 \sim N((1 - \xi)\mu, \sqrt{1 - \xi}\ \sigma) \tag{1a}$$

$$a_2 \sim N(\xi\mu, \sqrt{\xi}\ \sigma) \tag{1b}$$

$$a_1 + a_2 \sim N(\mu, \sigma) \tag{1c}$$

where a_1 - discrimination parameter for dimension 1 (major)
a_2 - discrimination parameter for dimension 2 (minor)
μ - mean of discrimination parameter for the whole test
σ - standard deviation of the a-parameter for the test
ξ - strength of major trait relative to the minor traits.

If ξ is 0.0, for example, then there are no minor traits, and the test data is strictly unidimensional. If ξ is 0.3, the potency of the major trait relative to the minor traits is about 70 percent. If ξ is set equal to 0.5, on the other hand, then the minor traits are not minor any more; their potency is equivalent to that of the major trait.

For this study, ξ values of 0.0, 0.3, and 0.5 were chosen. The first value (0.0) reflects a strictly unidimensional test which might be difficult to attain in practice. An ξ of 0.3 reflects test data with a dominant trait which affects all items and several notable minor traits (chosen to be 3 in this study) each affecting a cluster of items. In this setup, each test item will be influenced by the major trait and one of the minor traits. The third value of ξ (0.5) was chosen to reflect test data with one major trait that influences all items and several equally potent other traits specific to clusters of items. In this case, each item is influenced by two equally potent dimensions.

Based on earlier research studies (Ackerman, 1989; Ansley & Forsyth, 1985), three levels of correlations between the dominant ability and the minor traits were chosen. These were 0.0, 0.5, and 0.8.

Data Simulation. Four abilities from standard normal distribution were generated for 1000 examinees. The first ability (θ_m) was intended for the dominant trait whereas the other three abilities (θ_k, k=1,2,3) were intended for the minor traits. Three levels of correlation (ρ=0.0, 0.5, & 0.8) between the dominant ability (θ_m) and each of the minor abilities (θ_k) were examined. Difficulty values for 60 items were generated from a normal distribution N(-0.53,1). The item discrimination values were generated from a normal distribution with mean of 0.6 and standard deviation of 0.2. The choice of the item parameters was based on an analysis of mastery test data.

Each item discrimination index was resolved into two parts; discrimination along the major ability dimension and discrimination along the minor ability dimension. The partition of the discrimination value determined the strength of the minor traits relative to the major trait. This is only one of the several ways to model dimensionality of a test. For example it possible to have independent discrimination values along the major and minor traits. A bivariate extension of the two parameter logistic model (Reckase, 1985) was used to generate the test data. The model can be written as:

$$P_i = \frac{1}{1 + \exp\{-D[a_1(\theta_m - b_1) + a_2(\theta_k - b_2)]\}} \qquad (2)$$

where:

p_i is the probability of answering item i correctly
θ_m is the dominant ability
θ_k is the kth minor ability
D is an scaling factor equal to 1.7
a_1 is the discrimination of item i in the major dimension
a_2 is the discrimination of item i in the minor dimension
b_1 is the difficulty of item i in the major dimension
b_2 is the difficulty of item i in the minor dimension.

The probability of each examinee answering an item correctly was computed using equation 2. Uniform random numbers in the interval [0,1] were generated and compared with each examinee's probability. If the probability was less than the random number, the examinee was given a score of 0, and if the probability was greater than or equal to the generated random number, the examinee was given to a score of 1. This resulted in a 60-item binary test which has 20-item clusters and dimensionality controlled by the value of ξ. Note that each set of 20 successive items were influenced by one of the minor traits.

Data Analysis. At each level of ξ and $\rho(\theta_m, \theta_k)$, the simulated data were analyzed as follows:

1. The 60-item test was calibrated with the two-parameter logistic model using BILOG (Mislevy & Bock, 1990).
2. The mean of the standard errors of the ability estimates (SEE) was obtained from the BILOG output (ability score file).
3. Using the estimated item parameters, SEEs were computed for three ability scores: -1.0, 0.0, and 1.0.
4. Confidence intervals for 95% level were built around the three ability scores in step 3.
5. Each of the 20-item clusters (1-20, 21-40, 41-60) were calibrated with the two-parameter logistic model using BILOG.
6. Steps 2-4 above were repeated for each of the 20-item clusters
7. Fifty replications were performed in steps 1 through 6.

For the 60-item test, the means of the SEEs, the SEEs at the three ability scores, and the confidence intervals were compared across the three levels of ξ and the three levels of $\rho(\theta_m, \theta_k)$. Similar comparisons were made in the 20-item tests.

Results.

Table 1 summarizes the mean standard errors of the ability estimates (SEE) as obtained from the BILOG output. It also shows the correlations between the true (dominant) ability and the estimated ability (r), and the mean squared difference (MSD) of each pair. For brevity, results for only one of the three 20-item sets are presented (21-40). The results for all 20-item sets were similar. Cases where the major ability was correlated with the minor abilities and $\xi = 0$ were not examined as these would provide results similar to the first row of the table (see equation 2). When $\rho(\theta_m, \theta_k)$ was set at 0.0 and ξ was increased, the standard errors of the ability estimates (SEE) increased. As can be seen in the same rows, the values of r decreased, and the values of MSD increased with increase in ξ. These results are similar to findings from earlier studies (Ackerman, 1989; Ansley & Forsyth, 1985).

Insert Table 1 About Here

Recall that the 20-item clusters were purely unidimensional when $\rho(\theta_m, \theta_k)=0.0$ and $\xi=0$. In this case, values of 0.608, 0.899, and 0.327 were found for SEE, r, and MSD, respectively. At $\rho(\theta_m, \theta_k)=0.0$ and $\xi=0.5$, the 60-item tests resulted in values of 0.621, 0.778, and 0.433 for SEE, r, and MSD, respectively. Comparison of the two sets of values suggest that a 20-item unidimensional test may result in a better precision than a 60-item less unidimensional test in which the minor and major abilities are not related. This result is not surprising since a large portion of the precision expected from the 60-item test is not accounted for.

When $\rho(\theta_m, \theta_k)$ values were greater than 0.0 (major ability was correlated with the minor abilities), the MSD increased with increase in ξ, and r decreased with increase in ξ. Note that at both levels where $\rho(\theta_m, \theta_k)>0.0$, the differences in r and MSD between ξ values of 0.3 and 0.5 were less than the comparable differences at $\rho(\theta_m, \theta_k)=0.0$.

The SEEs at $\rho(\theta_m, \theta_k)>0.0$ portray a picture different from that found at $\rho(\theta_m, \theta_k)=0.0$. The SEE values decreased as ξ was increased at both levels of $\rho(\theta_m, \theta_k)>0.0$, and at both test lengths. Surprisingly, some of the SEEs at n=20 were even smaller than the SEEs found at the strictly unidimensional 20-item test. Although SEEs generally dropped as the major and minor traits were correlated, more precision was obtained when each item's discrimination value was divided equally between the major and the minor abilities. Since precision is related to the item discrimination values, one would expect less precision in estimating θ_k at $\xi=0.3$ than at $\xi=0.5$. This could lead to less precision at $\xi=0.3$ as compared to $\xi=0.5$, because the ability estimates from BILOG are best related to the average of θ_m and θ_k (see Ackerman, 1989).

In Table 2, the standard errors of estimates at selected ability levels are shown for $\rho(\theta_m, \theta_k)=0.0$. These standard errors were obtained by using item parameter estimates provided by the BILOG program. Ninety five percent confidence intervals which correspond to the standard errors are also shown. The standard

---

Insert Table 2 About Here

---

errors (and the confidence intervals) increased with increase in ξ. The pattern in Table 2 was apparently similar to that reported in the first three rows of Table 1. Again, unidimensional 20-item tests provided more precision than less unidimensional 60-item tests (at abilities 0 & 1) when the major trait was not related to the minor traits.

Table 3 presents the SEEs and confidence bands for $\rho(\theta_m, \theta_k)=0.5$. The standard errors were smaller, albeit marginally, when ξ was set at 0.5. To put it differently, more precision was obtained when each item equally discriminated along the major and minor traits. This provided smaller errors than putting more discrimination along the major trait. The differences between comparable SEE's was more notable at the 20-item test level. This could be explained by the fact that we have

----

Insert Table 3 About Here

----

larger number of abilities at the 60-item level than at the 20-item level. At the 60-item level, we are dealing with θ_m which affects all items and three minor θ's which affect 20 items each. At the 20-item level, on the other hand, we are dealing with θ_m which affects all 20 items and one θ_k which also affects all 20 items.

Similar results were obtained for $\rho(\theta_m, \theta_k)=0.8$ as shown in Table 4. Again the standard errors were smaller at $\xi=0.5$ than at $\xi=0.3$ (an exception was n=60 at 1). Dividing each item's discrimination value between the major ability and the minor ability resulted in less standard errors than when more of each

---

Insert Table 4 About Here

---

item's discrimination was allotted to the major ability. As in $\rho(\theta_m, \theta_k)=0.5$, the difference in SEEs between the two levels of ξ was larger at the 20-item level.

Between the two levels of $\rho(\theta_m, \theta_k)>0.0$, the 60-item tests resulted in almost equal SEE's (an average difference of 0.006), whereas the 20-item tests resulted in SEE's greater at $\rho(\theta_m, \theta_k)=0.8$ (an average difference of 0.042).

In almost all testing situations where IRT is used for scoring, it is normal to rescale the ability scores as well as the SEEs in order to present the examinee scores in a more interpretable way. In the State of Connecticut, for example, we rescale examinee ability scores so that the new scores would have a mean of 250 and an standard deviation of 45 (range of 100 to 400). If we use these scaling constant in the data presented in the preceding tables, the SEEs and their differences might be seen more clearly. Table 5 presents rescaled SEEs and related confidence intervals at an ability score of 0.0 for Tables 2, 3, and 4. At the 20-item level, an SEE difference of 3.1 is evident

---

Insert Table 5 About Here

---

for $\rho(\theta_m, \theta_k)=0.5$, and difference of 0.9 for $\rho(\theta_m, \theta_k)=0.8$. The differences are marginal at the 60-item level.

Figure 1 shows scaled mean error estimates at different levels of ξ and $\rho(\theta_m, \theta_k)$, and Figure 2 shows confidence bands around the mean scaled score.

Insert Figures 1 & 2 About Here

Summary and Conclusions.

The purpose of this study was to examine the impact of lack of unidimensionality on examinee IRT ability scores and test score reports. Three levels of degrees of dimensionality as controlled by the strength of the minor traits, and three levels of correlations between the major trait and each of the minor traits were simulated. The standard error of the ability estimates (SEEs) were compared among the different data sets. Another set of SEEs were computed from the item parameter estimates in order to compare measurement precision of the data sets in relation to the test items.

It was found that SEEs significantly increase with ξ when the major trait was not correlated with the minor traits. When the major and minor traits were correlated, however, the SEEs modestly decreased with increase in ξ. This could be attributed to the shift of more discriminating power toward the minor traits as ξ was increased. The decrease in SEEs as ξ was increased was more notable at the 20-item level. It was found that a 20-item unidimensional test could result in SEEs smaller than SEEs that would result from a 60-item less unidimensional test when the major and minor traits are not correlated.

Evidently, test dimensionality adversely affects the stability of the examinee scores if the trait which the test is purported to measure is not related to specific - unintended traits in the data. In these instances the stability of score reports may decline with a less unidimensional test. In fact, a more unidimensional subtest may produce more stable scores.

In the more realistic case where the intended and unintended traits are moderately correlated, increasing dimensionality by way of taking the specific traits into account may not affect the stability of test score reports. In fact, it might enhance the precision of the examinee score estimates if the reporting units are clustered around the specific traits.

To put this into practitioner's perspective, let us revisit the mathematics test example mentioned earlier in the paper. The test has different but related sections, say algebra, geometry, and statistics. The ability θ_m is the general mathematics proficiency, while three minor abilities θ_k are specific for each of the three clusters. The ability estimates by a unidimensional model would be the average of (θ_m, θ_k), and would be estimated best if examinees are adequately discriminated along both θ_m and θ_k. This is especially true if the scores are reported in clusters.

References

Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. Applied Psychological Measurement, 15, 13-24.

Ansley, T., & Forsyth, R. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. Applied Psychological Measurement, 9, 37-48.

Committee of AERA, APA, & NCME to Develop Standards. (1985). Standards for educational and psychological testing. Washington DC: American Psychological Association.

Drasgow, F., & Parsons, C. (1983). Application of unidimensional item response models to multidimensional data. Applied Psychological Measurement, 7, 189-199.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage Publications Inc.

Harrison, D. A. (1986). Robustness of IRT parameter estimation to violations of the unidimensionality assumption. Journal of Educational Statistics, 11, 91-115.

Mislevy, R., & Bock, D. (1990). BILOG: Maximum likelihood item analysis and test scoring with logistic models. Mooresville, IN: Scientific software.

Nandakumar, R. (1991). Traditional dimensionality versus essential dimensionality. Journal of Educational Measurement, 28, 99-117.

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

Shea, J., Norcini,J., & Webster, G. (1988). An application of item response theory to certifying examinations in internal medicine. Evaluation & the Health Professions, 11, 283-305.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. Psychometrika, 52, 589-618.

Traub, R. (1983). A priori consideration 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.

Table 1

Means of standard errors of estimates, and correlations and mean squared differences of estimated and true abilities

$\rho(\theta_m, \theta_k)$	Level of ξ	SEE n=60	SEE n=20	r n=60	r n=20	MSD n=60	MSD n=20
0.0	0.0	0.414	0.608	0.961	0.899	0.144	0.327
	0.3	0.533	0.704	0.905	0.779	0.215	0.457
	0.5	0.621	0.730	0.778	0.586	0.433	0.720
0.5	0.3	0.436	0.595	0.916	0.801	0.182	0.396
	0.5	0.425	0.543	0.856	0.699	0.295	0.566
0.8	0.3	0.427	0.614	0.952	0.877	0.118	0.283
	0.5	0.426	0.600	0.938	0.846	0.143	0.330

Table 2

Mean error estimates and confidence bands at selected ability levels for $\rho(\theta_m, \theta_k) = 0.0$

# of Items	Level of ξ	SEE at -1	SEE at 0	SEE at 1	Confidence band at -1	Confidence band at 0	Confidence band at 1
60	0.0	0.269	0.243	0.280	1.056	0.954	1.097
	0.3	0.347	0.355	0.414	1.360	1.391	1.621
	0.5	0.438	0.456	0.516	1.715	1.787	2.024
20	0.0	0.467	0.422	0.491	1.830	1.652	1.925
	0.3	0.554	0.567	0.681	2.171	2.222	2.671
	0.5	0.592	0.618	0.737	2.321	2.420	2.889

Table 3

Mean error estimates and confidence bands
at selected ability levels for
$\rho(\theta_m, \theta_k) = 0.5$

# of Items	Level of ξ	SEE at -1	SEE at 0	SEE at 1	Confidence band at -1	Confidence band at 0	Confidence band at 1
60	0.3	0.259	0.258	0.333	1.015	1.013	1.306
	0.5	0.252	0.248	0.324	0.988	0.972	1.270
20	0.3	0.406	0.401	0.550	1.590	1.570	2.155
	0.5	0.358	0.333	0.491	1.404	1.306	1.923

Table 4

Mean error estimates and confidence bands
at selected ability levels for
$\rho(\theta_m, \theta_k) = 0.8$

# of Items	Level of ξ	SEE at -1	SEE at 0	SEE at 1	Confidence band at -1	Confidence band at 0	Confidence band at 1
60	0.3	0.253	0.251	0.324	0.991	0.983	1.272
	0.5	0.250	0.249	0.327	0.981	0.976	1.282
20	0.3	0.432	0.428	0.560	1.692	1.676	2.195
	0.5	0.409	0.408	0.555	1.603	1.597	2.176

Table 5

Rescaled mean error estimates and confidence bands at ability = 0.0

# of Items	Level of ξ	SEE $\rho(\theta_m, \theta_k)$ 0.0	0.5	0.8	Confidence band $\rho(\theta_m, \theta_k)$ 0.0	0.5	0.8
	0.0	10.9	-	-	42.9	-	-
60	0.3	16.0	11.6	11.3	62.6	45.6	44.2
	0.5	20.5	11.2	11.2	80.4	43.7	43.9
	0.0	19.0	-	-	74.3	-	-
20	0.3	25.5	18.1	19.3	100.0	70.7	75.4
	0.5	27.8	15.0	18.4	108.9	58.8	71.9

Figure 1. Scaled Mean Error Estimates for the 20-item Level

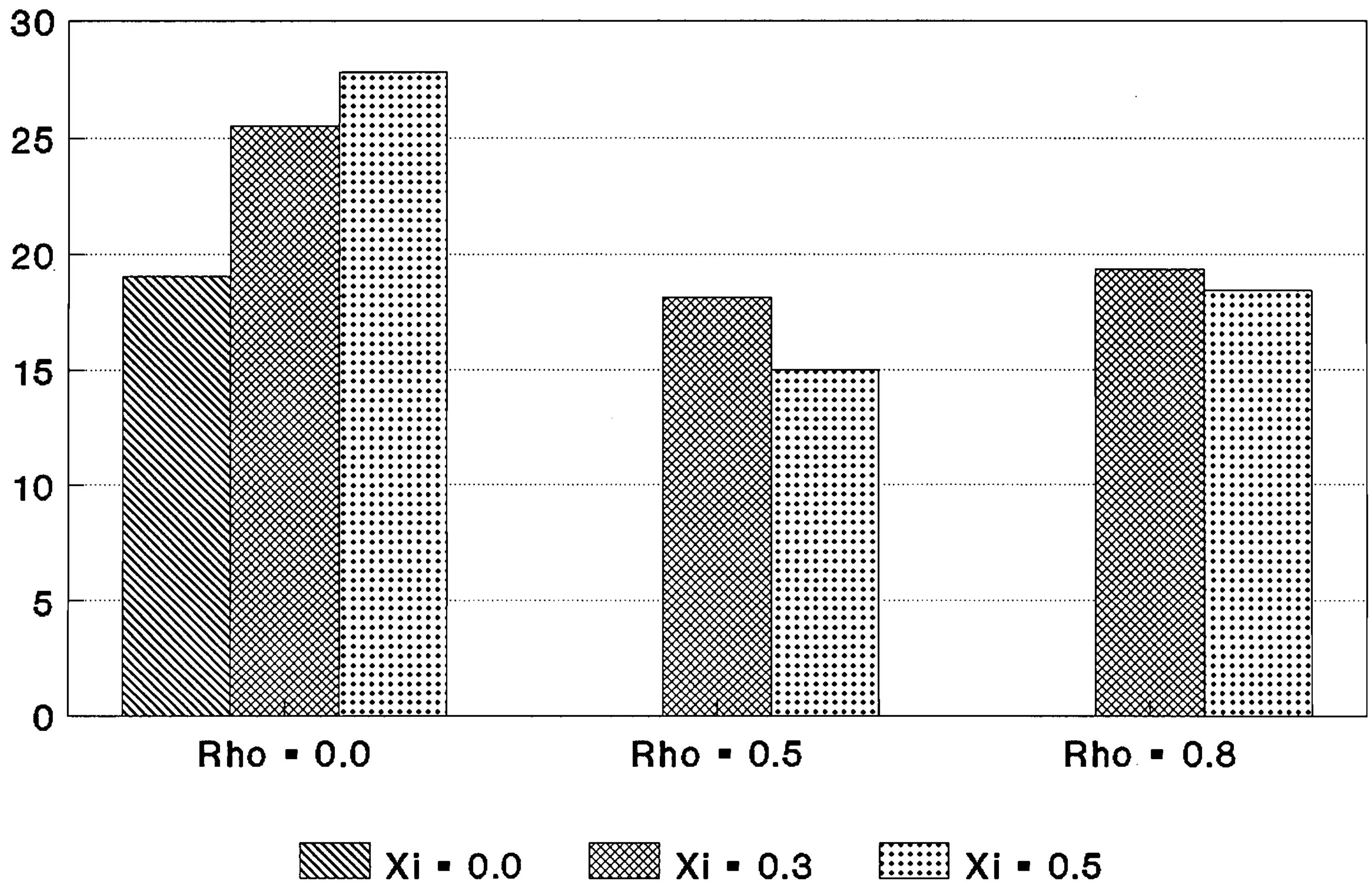
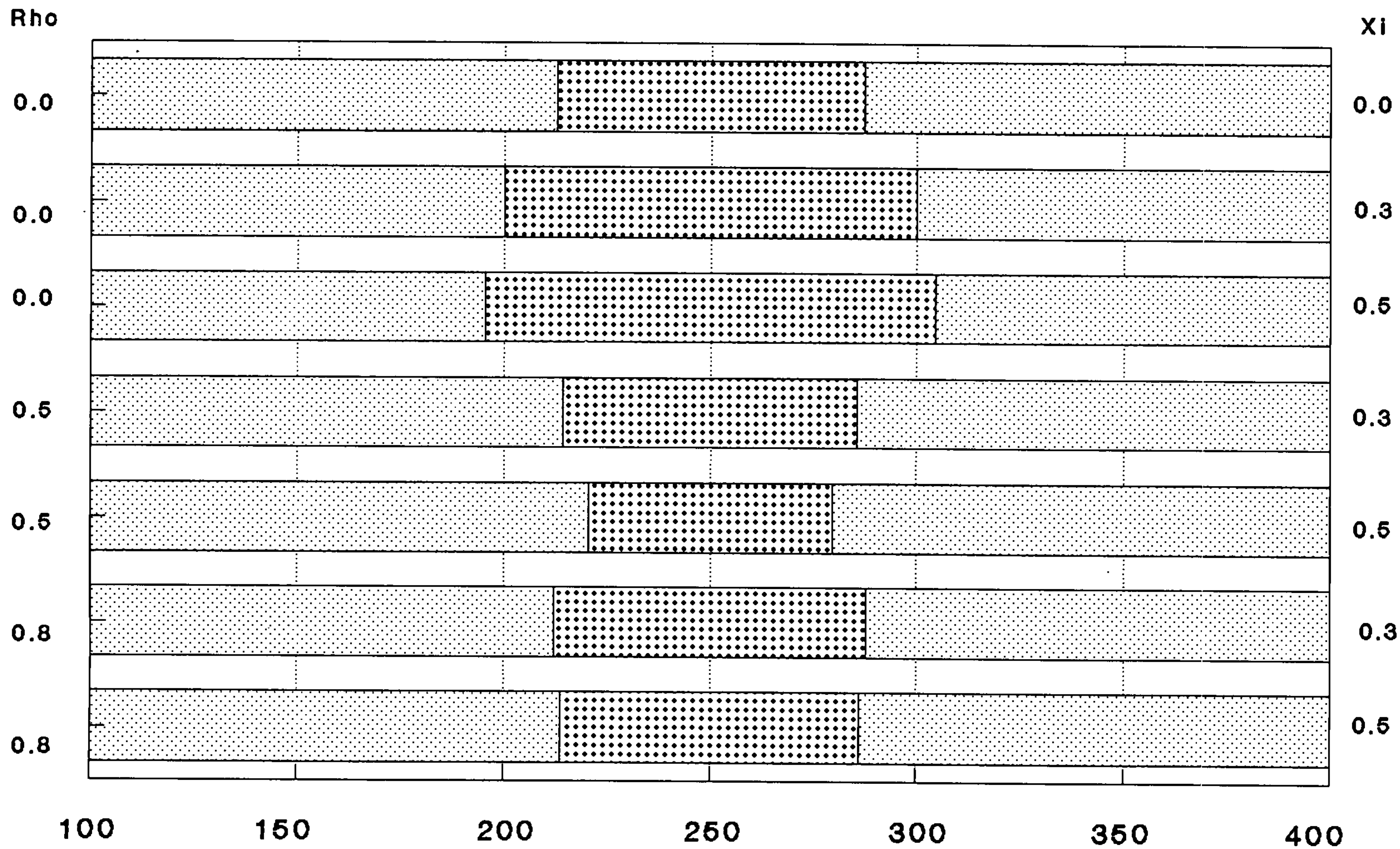


Figure 2. Confidence Bands for the Mean of the Scaled Scores (n=20)



TM 025806

NCME Annual Meeting, April 9-11, 1996

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: On Reporting IRT Ability Scores When the Test is not Unidimensional	
Author(s): Mohamed A. Dirir & Norma Sinclair	
Corporate Source:	Publication Date: April, 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

☑ Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document ☐

or here
Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Mohd	Position: Education Service Specialist
Printed Name: MOHAMED A. DIRIR	Organization: Connecticut Department of Educ.
Address: P.O. Box 2219 Hartford, CT 06145	Telephone Number: (860) 566-7156
	Date: May 8, 1996

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

March 12, 1996

Dear NCME Presenter,

Congratulations on being a presenter at NCME[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a written copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the NCME Conference. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (23)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: NCME 1996/ERIC Acquisitions
O'Boyle Hall, Room 210
The Catholic University of America
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the NCME web page (http://www.assessment.iupui.edu/ncme/ncme.html). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an NCME chair or discussant, please save this form for future use.

ERIC® Clearinghouse on Assessment and Evaluation