DOCUMENT RESUME

ED 052 241 TM 000 647

AUTHOR	Cahen, Leonard S.; And Others
TITLE	A Comparison of School Mean Achievement Scores With Two Estimates of the Same Scores Obtained by the Item-Sampling Technique.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RB-70-64
PUB DATE	Nov 70
NOTE	26p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Academic Achievement, Analysis of Variance, Comparative Analysis, *Comparative Testing, Correlation, *Grade 12, *Item Sampling, *Mathematics, Performance Factors, Statistics, Student Testing
IDENTIFIERS	Project TALENT Mathematics Test (Part 2), PTMT 2

ABSTRACT

The accuracy of estimating test means for groups of twelfth-grade students by the item-sampling technique was examined. The subjects were from 35 twelfth-grade schools participating in the National Longitudinal Study of Mathematical Abilities. Half of the students in each school were assigned to a treatment condition where they took a complete 24-item mathematics test on the first day of testing and took item-sampled versions of the same test on the second day of testing. A second random group of students within each of the schools took the item-sampled version of the mathematics test on day 2 but did not take the complete version of the mathematics test on day 1. There was no evidence to indicate that taking the complete 24-item mathematics test influenced the performance on the item-sampled version on the second day of testing. Reasonably close estimates of mean performance were obtained from the item-sampling situation as compared to the means estimated from the conventional type of testing. The differences between the means estimated from conventional type testing and from item-sampling testing were found to diminish as a function of the number of students tested in the school (square root transformation). (Author/PR)

RESEARCH BULLETIN

RB-70-64

A COMPARISON OF SCHOOL MEAN ACHIEVEMENT SCORES WITH TWO ESTIMATES OF THE SAME SCORES OBTAINED BY THE ITEM-SAMPLING TECHNIQUE

Leonard S. Cahen
Educational Testing Service

Thomas A. Romberg
University of Wisconsin

and

Walter Zwirner
University of Calgary

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the authors. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
November 1970

A COMPARISON OF SCHOOL MEAN ACHIEVEMENT SCORES WITH TWO ESTIMATES OF THE SAME SCORES OBTAINED BY THE ITEM-SAMPLING TECHNIQUE

Leonard S. Cahen
Educational Testing Service

Thomas A. Romberg
University of Wisconsin

and Walter Zwirner
University of Calgary

Abstract

The study examined the accuracy of estimating test means for groups of twelfth-grade students by the item-sampling technique. The subjects used in the study were from 35 twelfth-grade schools participating in the National Longitudinal Study of Mathematical Abilities. Half of the students in each school were assigned to a treatment condition where they took a complete 24-item mathematics test on the first day of testing and took item-sampled versions of the same test on the second day of testing. A second random group of students within each of the schools took the item-sampled version of the mathematics test on day 2 but did not take the complete version of the mathematics test on day 1. There was no evidence to indicate that the taking of the complete 24-item mathematics test influenced the performance on the item-sampled version on the second day of testing. Reasonably close estimates of mean performance were obtained from the item-sampling situation as compared to the means estimated from the conventional type of testing. The differences between the means estimated from conventional type testing and from item-sampling testing were found to diminish as a function of the number of students tested in the school (square root transformation).

A COMPARISON OF SCHOOL MEAN ACHIEVEMENT SCORES WITH TWO ESTIMATES OF THE SAME SCORES OBTAINED BY THE ITEM-SAMPLING TECHNIQUE[1]

Leonard S. Cahen
Educational Testing Service

Thomas A. Romberg
University of Wiscon^in

and Walter Zwirner
University of Cal.:ry

This study examines how accurately test means for groups of students can be estimated by the item-sampling technique. The statistical model for estimating means from item-sampled matrices has been outlined by Lord (1962) and by Lord and Novick (1968). Item sampling or matrix sampling refers to a general sampling procedure symbolized by a matrix with rows representing units (schools in this study) and columns defining measures. The measures are computed by Lord's (1962) procedures which involve the splitting of a set of <u>m</u> items into <u>k</u> random subsets of items. The <u>k</u> subsets are then assigned randomly to pupils in the <u>p</u> sampling units. The utility of this procedure is in its potential efficiency. Since each student in the sampling unit takes only a proportion of the items in the population of items, considerable administrative test time is saved. Cronbach (1963) has argued for the use of item-sampling in evaluation studies as an efficient way to estimate group means.

Item-sampling studies have been reported by Cahen, Romberg, and Zwirner (1970), Cook and Stufflebeam (1967), Jacobs and Wildemann (1969), Knapp (1968), Lord (1962), Osburn (1967), Owens and Stufflebeam (1969), Plumlee (1964), Shoemaker (1970), and Sirotnik (1970). With the exception of the studies by Cahen, Romberg, and Zwirner, Jacobs and Wildemann, Owens and Stufflebeam, and Sirotnik, these studies have used existing banks of data, i.e., samples were drawn post facto from existing test data banks in which each student had taken every item in a complete test version.

The present study is based on test data obtained under a priori item-sampling administration conditions. The test used in the study was the Project Talent Mathematics Test, Part 2 (PTMT-2) (Flanagan, Dailey, Shaycoft, Gorham, Orr, & Goldberg, 1962), a 24-item test containing fairly difficult mathematics items. The test was administered in the spring to twelfth-grade nonmathematics students participating in the National Longitudinal Study of Mathematical Abilities (NLSMA) (Cahen, 1965). The total NLSMA twelfth-grade testing population was divided into mathematics and nonmathematics classifications (Populations 1 and 2) (Romberg & Wilson, 1969). Population 2 NLSMA students, the ones used in this study, were defined as those students not completing a full year of mathematics beyond algebra and geometry. Population 1 students had taken a mathematics course beyond algebra and geometry and were considered superior mathematics students compared to NLSMA Population 2 students.

For the item-sampling study, a random half of the Population 2 students in each sampling unit (the school) first took all of the items in the PTMT-2. The following day the same students took an item-sampled version of the same 24-item test. This allowed a comparison between the school means obtained on the complete version of the test on day one (Measure A) with the estimate of the mean obtained by item sampling for the same students on day two (Measure B). One additional mean (Measure C) was obtained by the item-sampling technique and was computed for a group of Population 2 subjects in each school who had not taken the complete PTMT-2 on the previous day. Figure 1 illustrates the testing schema.

Insert Figure 1 about here

A more sophisticated approach would have used a design controlling for order of taking the complete 24-item test and the item-sampling version. However, the more elaborate design was not feasible within the testing program of NLSMA.

The complete 24-item test was administered as a separately timed 24-minute section in the NLSMA Spring Battery. Students thus had an average of one minute working time per item. For the item-sampled form, the PTMT-2 items were embedded in 15 item-sampled booklets or forms. The population of sampled items totaled 120 items from NLSMA Test Forms 122 A, B, and C, a group of mathematics achievement items that had been administered to the same students two and a half years earlier when the students were tenth-grade NLSMA participants. Students were also allowed an average of one minute working time per item on the test items from NLSMA Test Forms 122 A, B, and C.

Table 1 shows the item placement in the 15 item-sampled booklets.

Insert Table 1 about here

The PTMT-2 items were randomly assigned without replacement to the 15 booklets or forms. Each booklet received either one or two of the total 24 PTMT-2 items and either six or seven NLSMA items, making a total of eight mathematics items per item-sampling booklet. The students were allowed a total of eight minutes to complete the eight item-sampling items. In this study, therefore, the students were allowed the same amount of working time on the item-sampling administrations as they were allowed on the administration of the complete PTMT-2. This hopefully overcame one of the deficiencies in the study reported by Cahen, Romberg, and Zwirner (1970) where students potentially had more time per item on the item-sampled testing condition than on the conventional testing condition. Within each booklet the eight items were then randomly assigned to the eight item positions. Finally, the PTMT-2 items were rearranged to preserve the order in which they appeared in the regular 24-item test. The NLSMA items from Forms 122 A, B, and C were also rearranged so that they appeared in the same serial order as they would appear in the complete NLSMA 122 A, B, and C tests.

After printing, the booklets were collated by sets or blocks of order permutations of 15. One copy of each of the 15 booklets was placed in one permuted order, a second set of 15 booklets was placed in a different permuted order of 15, etc. The teachers were told to pass out the booklets so that the first student received the top booklet, the second student received the second booklet, etc. The distribution of booklets to pupils therefore approximated a random assignment.

In the spring testing session there were 149 schools with twelfth-grade students participating in NLSMA. However, in most of these schools, very few students participating in the testing program were identified as the non-mathematics students (Population 2 students). In order to estimate a school mean for these students, it was decided that data must be available for a minimum of five students on both the complete PTMT-2 and on the item-sampled version of the PTMT-2. This decision is admittedly arbitrary. A more restrictive criterion would have required that each PTMT-2 item be attempted by at least one student in each school. However, using the minimum requirement of five students, school means (Measures A, B, and C) were calculated for 35 schools.

Insert Table 2 about here

Within each school, the number of students for Measures A and B are the same as these students took the complete test on day one and the item-sampling version on day two.

Table 3 shows the means, standard deviations, and intercorrelations of the school means on Measures A, B, and C. The difference between the AB

Insert Table 3 about here

----

correlation and the AC correlation was tested by the procedure developed by Olkin (1967, pp. 111-113) to test for differences between correlated coefficients of correlation. A 95 per cent confidence band was established around the difference between the AB and AC coefficients. The value of zero barely fell in the confidence interval for the difference between the correlation coefficients for AB and AC ($-.012 \leqq r_{AB} - r_{AC} \leqq .134$). It was concluded that the correlation coefficients of AB and AC were not necessarily significantly different. It should be pointed out that the correlation coefficient AB was computed from measures obtained on the same students. This is not the case for the correlation between Measures AC. Measures A and B were obtained for one set of students within a school while Measure C was obtained from a random but different subset of students within the same school. The correlation AB might be interpreted as a test-retest coefficient while the AC coefficient represents the association between measures obtained on two random subset samples from a sampling unit. The additional information about reliability was obtained by computing two intraclass coefficients of correlation from the data displayed in Table 2. Using the 35 schools representing rows and the three measures (A, B, and C) representing columns, an intraclass correlation of .835 was obtained. When the two measures estimated by item sampling (B and C) form the columns of the matrix, a coefficient of .679 was obtained.

The data matrix displayed in Table 2 was also submitted to a multivariate analysis of variance.[2] The 35 rows representing schools in Table 2 were stratified into three levels defined by the ranges of number of students tested in the school. The dependent measures for the analysis were A, B, and C. Table 4 provides data about the ranges of the number of students tested in each school by each stratum, number of schools in these defined analysis units, and the means and standard deviations for Measures A, B, and C by strata.

Insert Tables 4 and 5 about here

Table 5 shows the results of the multivariate analysis of variance from the schema displayed in Table 4. The dimensions of measures and the interaction of measures by strata yielded F values of less than one. The stratification dimension yielded an F value of 2.84 ($p < .063$). On the average the schools with the smaller number of students tested yielded higher mean scores on Measures A, B, and C than the stratum two and stratum three schools as defined in Table 4. On the average, the mean performance on Measures A, B, and C was lower for the stratum three schools than for the stratum two schools.

It appears that the item-sampling procedure provides reasonably precise estimates of school achievement. This statement is supported by the correlational data reported in Table 3 and by the small variance attributed to the dimension of measures as shown in Table 4. Table 6 provides additional information about the magnitude of the difference scores between Measures AB, AC, and BC.

Insert Table 6 about here

Another point must be considered in the interpretation of the magnitude of the difference scores shown in Table 6. Cahen, Romberg, and Zwirner (1970), in an earlier paper on the item-sampling procedures, pointed out that errors of measurement are present in both measures obtained from conventional sampling as well as from scores obtained from item-sampling. These errors of measurement will attenuate the measures of association between Measures A, B, and C. In the previous research reported by Cahen, Romberg, and Zwirner (1970), it was shown that the magnitude of error between the means estimated by conventional sampling and from item sampling decreased as the number of students tested in the school increased. They reported that the error between these two measures decreased as a function of the square root of the number of students tested.

In order to study the relationship between magnitude of error as a function of the number of students tested in the school from the data available here, a fourth measure (Measure D) was created. Measure D is a weighted mean of Measures B and C. The estimate of this combined mean was computed by a formula provided by McNemar (1962, p. 18). Data on Measure D, the number of students tested in each school used in estimating Measure D, and the deviation of Measure D from Measure A, are shown in Table 7. The n for Measure D is the sum of the n's for Measures B and C.

Insert Table 7 about here

The signs (plus or minus) of the deviation of Measure D from Measure A showed an approximately equal number of pluses and minuses. Also, there was no statistical relationship between the number of students tested in the school and the sign of the deviation of Measure D from Measure A. It was therefore decided to use the absolute deviation of Measure D from Measure A in studying the relationship of this measure to the number of students tested in the school.

Table 8 shows a bivariate plot of the regression of the absolute deviation between Measures D and A on the number of students tested in the school. A square root transformation was used on the latter variable. The data

----

Insert Table 8 about here

----

displayed in Table 8 show a decrease in the magnitude of the absolute deviation of Measure D as the number of students tested in the school increases.

The correlation between Measure D and the square root of the number of students tested was .398 which is significant at the .05 level.

In addition to fitting the data by a simple linear regression model, a second regression analysis was performed by adding a second degree polynomial term. The addition of the second degree polynomial term into the regression equation did not significantly provide a better fit of the regression of the absolute deviation on the number of students tested.

We believe this study provides additional empirical information on the practical utility of the item-sampling procedure. As in the previous study reported by Cahen, Romberg, and Zwirner (1970), it has been shown that the measurement error attributable to item sampling decreases as the number of students in the schools increase.

The researcher-evaluator who intends to use the item-sampling procedure is again encouraged to consider the magnitude of error he can afford in his estimation of estimates of means for sampling units (classes, schools, etc.). With this type of decision making strategy it should be possible for the researcher-evaluator to determine the minimum number of students he needs in his sampling units in order to preserve a reasonable magnitude of measurement error.

References

Bock, R. D. Programming univariate and multivariate analysis of variance. Technometrics, 1963, 5, 95-117.

Cahen, L. S. An interim report on the national longitudinal study of mathematical abilities. Mathematics Teacher, 1965, 58, 522-526.

Cahen, L. S., Romberg, T. A., & Zwirner, W. The estimation of mean achievement scores for schools by the item-sampling technique. Educational and Psychological Measurement, 1970, 30, 41-60.

Clyde, D. J., Cramer, E. M., & Sherin, R. J. Multivariate statistical programs. Biometric Laboratory, University of Miami, Coral Gables, Fla., 1966.

Cook, D. L., & Stufflebeam, D. L. Estimating test norms from variable size item and examinee samples. Educational and Psychological Measurement, 1967, 27, 601-610.

Cronbach, L. J. Evaluation for course improvement. Teachers College Record, 1963, 64, 672-683.

Flanagan, J. C., Dailey, J. T., Shaycoft, M. F., Gorham, W. A., Orr, D. B., & Goldberg, I. Design for a study of American youth. Boston: Houghton-Mifflin, 1962.

Graybill, F. A. An introduction to linear statistical models, Vol. 1. New York: McGraw-Hill, 1961. Pp. 223-253.

Jacobs, S. S., & Wildemann, C. Matrix sampling with classroom tests: An empirical investigation. Paper presented at the ERANYS Convocation, Concord Hotel, Kiamesha Lake, New York, November 1969.

Knapp, T. R. An application of balanced incomplete block designs to the estimation of test norms. Educational and Psychological Measurement, 1968, 28, 265-272.

Lord, F. M. Estimating norms by item-sampling. Educational and Psychological Measurement, 1962, 22, 259-267.

Lord, F. M., & Novick, M. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

McNemar, Q. Psychological statistics. (3rd ed.) New York: Wiley, 1962.

Olkin, I. Correlations revisited. In J. Stanley (Ed.), Improving experimental design and statistical analysis. Seventh Annual Phi Delta Kappa Symposium on Educational Research. Chicago: Rand-McNally, 1967. Pp. 102-128.

Osburn, H. G. A note on design of test experiments. Educational and Psychological Measurement, 1967, 27, 797-802.

Owens, T. R., & Stufflebeam, D. L. An experimental comparison of item sampling and examinee sampling for estimating test norms. Journal of Educational Measurement, 1969, 6, 75-83.

Plumlee, L. B. Estimating means and standard deviations from partial data--an empirical check on Lord's item-sampling technique. Educational and Psychological Measurement, 1964, 24, 623-630.

Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952. Pp. 236-272.

Romberg, T. A., & Wilson, J. W. The development of tests. In J. W. Wilson, L. S. Cahen, & E. G. Begle (Eds.), NLSMA Reports, No. 7, School Mathematics Study Group, Stanford University, Stanford, California, 1969.

Roy, S. N. Some aspects of multivariate analysis. New York: Wiley, 1957. Pp. 76-86.

Shoemaker, D. M. Allocation of items and examinees in estimating a norm distribution by item-sampling. Journal of Educational Measurement, 1970, 7, 123-128.

Sirotnik, K. An investigation of the context effect in matrix sampling. Journal of Educational Measurement, 1970, 7, 199-207.

Footnotes

[1]This research was supported in part by the National Institute of Child Health and Human Development, under Research Grant 1 P01 HD01762.

The authors are indebted to Frederic M. Lord for his suggestions about applications of item-sampling procedures in the National Longitudinal Study of Mathematical Abilities, and for his comments on an earlier draft of this paper. Appreciation is also extended to Robert L. Linn for critically reviewing an earlier draft of this paper. Of course, they are not responsible for any errors or misinterpretations that remain. Data analyses were helpfully performed by John J. Ferris and William Van Hassel of the Office of Computation Sciences, Educational Testing Service.

The authors thank the School Mathematics Study Group for permission to use data collected for the National Longitudinal Study of Mathematical Abilities; however, the opinions stated in this paper are solely those of the authors.

[2]The multivariate analysis of variance was performed at Educational Testing Service on a program developed by Clyde, Cramer, and Sherin (1966) at the Biometric Laboratory, University of Miami. The technical manual states that the methods used in the MANOVA program come from the work of Bock (1963), Graybill (1961), Rao (1952), and Roy (1957).

Table 1

Item Source Key for NLSMA Year III Spring Battery Booklets

Booklets 12_{3A} through 12_{30} (Item-sampled)

12-3A

Item	Source	Item
7	122A	2
8	122A	17
9	122A	27
10	122A	37
11	PTMT-2	31
12	122B	23
13	122B	40
14	PTMT-2	37

12-3B

Item	Source	Item
7	122A	23
8	122A	30
9	PTMT-2	21
10	PTMT-2	36
11	122A	44
12	122B	1
13	122B	5
14	122B	19

12-3C

Item	Source	Item
7	122A	4
8	PTMT-2	40
9	122A	22
10	122A	26
11	122A	29
12	122B	4
13	122B	24
14	122C	24

12-3D

Item	Source	Item
7	122A	1
8	122A	21
9	122B	17
10	122B	21
11	122B	26
12	PTMT-2	20
13	122B	41
14	122C	37

12-3E

Item	Source	Item
7	122A	10
8	PTMT-2	28
9	122A	24
10	122A	43
11	122B	6
12	PTMT-2	38
13	122B	15
14	122B	44

12-3F

Item	Source	Item
7	122A	19
8	122A	25
9	122A	36
10	PTMT-2	19
11	122A	42
12	122B	39
13	122B	43
14	PTMT-2	39

SOURCE: NLSMA

SOURCE: 122A, 122B, and 122C (NLSMA Forms 122A, B, and C)

PTMT-2 (Project Talent Mathematics Test-Part 2)

Table 1 (cont'd)

12-3G Item	Source	Item	12-3H Item	Source	Item	12-3I Item	Source	Item
7	122A	8	7	122A	5	7	122A	6
8	122A	14	8	122A	12	8	122A	33
9	122A	15	9	122B	3	9	122A	40
10	122A	20	10	PTMT-2	24	10	122B	11
11	122B	7	11	PTMT-2	32	11	PTMT-2	18
12	122B	27	12	122B	9	12	122B	28
13	PTMT-2	26	13	122B	34	13	122B	35
14	122C	31	14	122B	36	14	122C	4

12-3J Item	Source	Item	12-3K Item	Source	Item	12-3L Item	Source	Item
7	122A	3	7	122A	9	7	PTMT-2	23
8	PTMT-2	30	8	122A	18	8	122A	7
9	PTMT-2	34	9	PTMT-2	29	9	PTMT-2	33
10	122A	11	10	122A	45	10	122A	16
11	122A	35	11	122B	18	11	122A	28
12	122A	39	12	122B	20	12	122A	32
13	122B	8	13	122B	33	13	122A	34
14	122B	37	14	122C	27	14	122A	41

Table 1 (cont'd)

12-3M			12-3N			12-30		
Item	Source	Item	Item	Source	Item	Item	Source	Item
7	122B	2	7	PTMT-2	27	7	122A	13
8	122B	13	8	122A	38	8	PTMT-2	22
9	PTMT-2	17	9	PTMT-2	35	9	122A	31
10	122B	16	10	122B	10	10	122B	25
11	122B	30	11	122B	12	11	122B	29
12	122B	31	12	122B	14	12	122B	32
13	PTMT-2	25	13	122B	22	13	122B	38
14	122B	42	14	122B	45	14	122C	10

Table 2

Measures A, B, and C for 35 Schools with Group 2 Students

in the NLSMA 12th Grade Population

School	N	Measure A	N	Measure B	N	Measure C
1	55	10.65	55	11.12	65	10.06
2	75	9.85	75	9.65	63	11.07
3	99	8.74	99	9.72	104	8.96
4	53	7.62	53	8.46	51	9.16
5	43	7.79	43	6.54	43	7.77
6	36	13.22	36	11.02	35	15.17
7	25	12.64	25	12.00	22	8.93
8	53	10.90	53	11.34	50	11.50
9	58	10.14	58	9.82	56	8.74
10	29	7.48	29	9.38	31	8.00
11	19	8.26	19	11.20	25	10.26
12	33	14.15	33	13.54	27	13.25
13	19	14.21	19	15.63	16	14.36
14	23	13.39	23	12.62	18	13.00
15	15	12.27	15	12.22	22	14.17
16	18	13.56	18	13.60	23	16.62
17	15	10.60	15	13.80	16	10.91
18	15	11.60	15	12.60	33	11.22
19	15	10.07	15	11.11	15	10.50
20	27	11.44	27	13.88	24	12.00
21	23	9.39	23	9.00	21	9.70
22	28	9.07	28	6.43	26	11.54
23	24	11.25	24	8.75	20	12.82
24	17	10.23	17	12.00	20	6.00
25	18	6.22	18	8.18	23	5.54
26	19	11.42	19	11.09	22	13.70
27	19	10.47	19	13.00	17	12.00
28	35	11.74	35	12.46	35	14.42
29	17	11.76	17	10.25	7	5.50
30	15	11.13	15	12.75	12	12.50
31	19	7.05	19	5.60	13	5.34
32	16	12.75	16	11.34	11	12.75
33	7	14.00	7	16.00	15	16.00
34	5	9.60	5	16.00	21	7.69
35	14	14.14	14	12.00	17	13.50

Table 3

Means, Standard Deviations, and Intercorrelations

for Measures A, B, and C

(N = 35)

	A	B	C	$\bar{X}$	σ
A	1.000	.688	.749	10.82	2.18
B		1.000	.521	11.26	2.54
C			1.000	10.99	3.01

Table 4

Means and Standard Deviations for Measures A, B, and C for 35 Schools Stratified by the Number of Students Tested in the Schools

Strata	Ranges of Number of Students Tested in Schools in Strata	N of Schools	Measure A	Measure B	Measure C
1	2 - 18	13	$\bar{X}$ 11.38 σ 2.15	$\bar{X}$ 12.45 σ 2.14	$\bar{X}$ 10.99 σ 3.81
2	19 - 50	16	$\bar{X}$ 10.81 σ 2.40	$\bar{X}$ 10.76 σ 2.91	$\bar{X}$ 11.39 σ 2.80
3	51 - 99	6	$\bar{X}$ 9.65 σ 1.25	$\bar{X}$ 10.02 σ 1.06	$\bar{X}$ 9.92 σ 1.16

Table 5

Multivariate Analysis of Variance of Measures A, B, and C for 35 Schools Stratified by the Number of Students Tested in the Schools

Source	SS	DF	MS	F	P
Measures	3.41	2	1.70	.26	.774
Strata	31.67	2	18.84	2.84	.063
Measures X Strata	15.80	4	3.95	.60	.666
Error	636.23	96	6.63		

Table 6

Number of Students Tested and Difference Scores between

Measures A, B, and C for 35 Schools

School	NA	NB	NC	A-B	A-C	B-C
1	55	55	65	-.47	.59	1.06
2	75	75	63	.20	-1.22	-1.42
3	99	99	104	-.98	-.22	.76
4	53	53	51	-.84	-1.54	-.70
5	43	43	43	1.25	.02	-1.23
6	36	36	35	2.20	-1.95	-4.15
7	25	25	22	.64	3.71	3.07
8	53	53	50	-.44	-.60	-.16
9	58	58	56	.32	1.40	1.08
10	29	29	31	-1.90	-.52	1.38
11	19	19	25	-2.94	-2.00	.94
12	33	33	27	.61	.90	.29
13	19	19	16	-1.42	-.15	1.27
14	23	23	18	.77	.39	-.38
15	15	15	22	.05	-1.90	-1.95
16	18	18	23	-.04	-3.06	-3.02
17	15	15	16	-3.20	-.31	2.89
18	15	15	33	-1.00	.38	1.38
19	15	15	15	-1.04	-.43	.61
20	27	27	24	-2.44	-.56	1.88
21	23	23	21	.39	-.31	-.70
22	28	28	26	2.64	-2.47	-5.11
23	24	24	20	2.50	-1.57	-4.07
24	17	17	20	-1.77	4.23	6.00
25	18	18	23	-1.96	.68	2.64
26	19	19	22	.33	-2.28	-2.61
27	19	19	17	-2.53	-1.53	1.00
28	35	35	35	-.72	-2.68	-1.96
29	17	17	7	1.51	6.26	4.75
30	15	15	12	-1.62	-1.37	.25
31	19	19	13	1.45	1.71	.26
32	16	16	11	1.41	0	-1.41
33	7	7	15	-2.00	-2.00	0
34	5	5	21	-6.40	1.91	8.31
35	14	14	17	2.14	.64	-1.50
$\bar{X}$				-0.44	-0.17	0.27
σ				1.89	2.00	2.75

Table 7

N, Measure D (Weighted Mean of Means B and C), and the Discrepancy between Measure D and Measure A

School	N of Measure D (Combined n's of Weighted Mean B & C)	Measure D (Weighted Mean of B & C Means)	Deviation of Measure D from Measure A
1	120	10.54	0 11
2	138	10.30	-0.45
3	203	9.33	-.59
4	104	8.80	-1.18
5	86	7.16	.63
6	71	13.06	.16
7	47	10.56	2.08
8	103	11.42	-.52
9	114	9.29	.85
10	60	8.67	-1.19
11	44	10.66	-2.40
12	60	13.41	.74
13	35	15.05	-.84
14	41	12.79	.60
15	37	13.38	-1.11
16	41	15.29	-1.73
17	31	12.31	-1.71
18	48	11.65	-.05
19	30	10.80	-.73
20	51	13.00	-1.56
21	44	9.33	.06
22	54	8.89	.18
23	44	10.60	.65
24	37	8.76	1.47
25	41	6.70	-.48
26	41	12.49	-1.07
27	36	12.53	-2.06
28	70	13.44	-1.70
29	24	8.86	2.90
30	27	12.64	-1.51
31	32	5.49	1.56
32	27	11.91	.84
33	22	16.00	-2.00
34	26	9.29	.31
35	31	12.82	1.32

Table 8

Bivariate Relationship between Absolute Deviation of Measure D and Measure A and the Number of Students Tested in the School (Square Root Transformation)

Absolute Deviation of Measures D and A	3.00 3.99	4.00 4.99	5.00 5.99	6.00 6.99	7.00 7.99	8.00 8.99	9.00 9.99	10.00 10.99	11.00 11.99	12.00 12.99	13.00 13.99	14.00 14.99
2.75 - 2.99		1										
2.50 - 2.74												
2.25 - 2.49				1								
2.00 - 2.24		1		2								
1.75 - 1.99												
1.50 - 1.74			3	1	1	1						
1.25 - 1.49			1	1								
1.00 - 1.24				2	1			1				
.75 - .99			3					1				
.50 - .74				2	1		1	1				1
.25 - .49			1	1					1			
.00 - .24				2	1	1		1				

Number of Students Tested
(Square Root Transformation)

Figure 1

Testing Schema

	Day 1	Day 2
Random Half One	Complete PTMT-2 (Measure A)	Item Sample PTMT-2 (Measure B)
Random Half Two	Different Test	Item Sample PTMT-2 (Measure C)