73-14,385

SPILMAN, Helen W., 1938-
THE USE OF A SINGLE ITEM-SAMPLE TO ESTIMATE GROUP ACHIEVEMENT.

The City University of New York, Ph.D., 1973
Education, psychology

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE USE OF A SINGLE ITEM-SAMPLE TO
ESTIMATE GROUP ACHIEVEMENT

by

HELEN SPILMAN

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1973

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

January 23, 1973
date

Max Weiner
Chairman of Examining Committee

January 23, 1973
date

Max Weiner
Executive Officer

Patricia Kay

Alan Gross

Max Weiner

Supervisory Committee

The City University of New York

Abstract

THE USE OF A SINGLE ITEM-SAMPLE TO ESTIMATE
GROUP ACHIEVEMENT

by

Helen Spilman

Adviser: Professor Max Weiner

The primary objective of this study was to investigate the use of a single sample of items drawn at random from a standardized test to provide estimations of a population mean and standard deviation on a total test. The effect of context on the performance of examinees on items-samples was an important consideration in this investigation.

Stratified random samples of 25 per cent of the total test were drawn from the 63 item Stanford Achievement Test, Test 6: Arithmetic (Stanford Arithmetic Test), Primary I (Forms X and Y) to provide one mini-test for each form. The subjects were 153 second grade pupils. All of the pupils were given the total Stanford Arithmetic Test (Form X). Two weeks later the subjects were divided into four groups for the administration of mini-tests. Two groups of pupils were administered mini-tests only: Group A received mini-test, Form X, and Group B was given mini-test, Form Y.

Two groups were administered the mini-tests as a part of a longer, more complex teacher-made arithmetic test: Group C received mini-test, Form X, and Group D was administered mini-test, Form Y.

Comparisons were made of parameter estimates derived from the two empirical (a priori) sampling administration conditions and of post hoc mini-test parameter estimates obtained using the item data from the administration of the total Stanford Arithmetic Test (Form X). These comparisons were made to study possible effects on pupil performance when items were administered in different contexts.

Both groups of subjects taking Stanford Arithmetic mini-tests as a part of the teacher-made test (Groups C and D) underestimated the group mean on the complete test. The two groups taking only a mini-test (Groups A and B) overestimated the actual group means.

None of the estimated means derived from the a priori mini-tests were within one standard error of measurement (SEM) of the total test means for the Stanford Arithmetic Test. The post hoc mini-test estimates were all within one SEM.

The a priori mini-test administration design was repeated using 235 second grade pupils and 25 per cent item-samples drawn from the Metropolitan Achievement Test in Reading, Primary I and II, Form F. Results were similar to those obtained from the Stanford Arithmetic Test mini-test administration.

To provide information on the effectiveness of the item-sampling approach used in this study, 100 twenty-five per cent stratified random item-samples were drawn from the complete Stanford Arithmetic Test. Estimated means, standard deviations and KR21 reliability coefficients for the total test were determined for each item-sample. The frequency distributions for the estimated parameters were plotted and the means and standard deviations, or the standard errors of estimate for the sampling plan used in this study, were determined for each distribution.

The actual mean, standard deviation and KR21 coefficient for the total test were 45.32, 11.65 and .93 respectively. The mean of the distribution of estimated means was 45.39, the mean of the estimated standard deviations was 11.59, and the mean of estimated KR21 coefficients was .92. The standard errors of estimate for the three parameters were 2.07 for the mean, .85 for the standard deviation and .03 for the KR21 coefficient. These results indicated that the single item-sample approach was efficient.

The use of a single sample of items for the estimation of group performance may prove to be an effective approach to item-sampling and more practical than the traditional multiple matrix-sampling approach. The general assumption that performance on item-samples is not affected by their context has been placed in question.

Further research is recommended: 1) to provide more information about the use of a single sample of items for estimating group parameters, and 2) to shed more light upon the issue of context effect.

ACKNOWLEDGEMENTS

The author wishes to express her deepest thanks to Dr. Carol Tittle for her unceasing willingness and good cheer in assisting her in so many ways from the conception of this project to its completion.

I am also grateful for the guidance and encouragement provided by the members of my dissertation research committee, Professor Max Weiner, Chairman, Professor Patricia Kay and Professor Alan Gross.

I am indebted to the second grade teachers and pupils from C.S. 234 and C.S. 67 for their cooperation in the testing required for this project.

My deep appreciation and gratitude go to my husband, Edra Spilman, for his moral support and assistance through the past three and a half years of study, trials and tribulations.

Finally, I also wish to thank Professor Donald Rock and Professor Edward Frankel for consenting to serve on my examination committee.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

LIST OF TABLES

LIST OF ILLUSTRATIONS

I. INTRODUCTION

An important problem in the use of standardized tests is the amount of time involved in their administration. A solution to this problem has been the use of item-samples rather than complete tests. This study investigated the following problems in the use of item-sampling for the evaluation of group performance:

1. The use of a single sample of items to estimate group achievement.
2. The effect of context on the performance of examinees on item-samples.
3. The evaluation of the "goodness" of an item-sampling procedure.
4. The reliability of item-sampling estimates of group performance.

During the past ten years, the interest in item-sampling for the estimation of group performance on achievement tests has grown rapidly as a result of research in this field. Most item-sampling has taken the form of matrix-sampling, i.e., the simultaneous random sampling of both subjects and items. A single matrix-sample consists of a sample of subjects taking a sample of items. The multiple matrix technique involves the use of several samples of subjects taking different samples of items.

Item-sampling has the advantage of improving efficiency in the testing situation through a decrease in testing time. It shows particular promise in the assessment of large groups of pupils. Item-sampling should prove to be a useful tool in the evaluation of instructional programs for schools, school districts and states.

One of the problems with the use of item-samples is the limited information currently available concerning the context effect on estimates of achievement. A "context effect" is defined as "...any tendency for the matrix-sample estimates of matrix population parameters to be different from those obtained had the examinees in fact responded to the population of items" (Sirotnik, 1970). For example, a test of 50 items and a test of 5 items are each presented in a different context which may possibly affect performance. Some of the factors which may play a role in the affect of context on test performance are practice, length of test (fatigue), motivational variables and order of item difficulty. Before item-sampling is recommended for general use in the estimation of group performance, the factors involved in context effects on performance should be more carefully examined.

The administration of multiple matrix-samples introduces problems in logistics not present in the use of a single test or item-sample: a number of different tests must be constructed and reproduced, students must be distributed into random groups with each group receiving a test, and the several tests must be administered and scored.

In contrast to the more traditional matrix-sampling or to examinee-sampling, where an entire test is administered to a random sample of examinees, the use of a single item-sample to estimate group parameters on a test would have distinct practical advantages. In terms of the logistics of test administration, the use of a single item-sample would be more efficient and less arduous than the use of multiple samples. A single item-sample, or mini-test, would be constructed and reproduced rather than multiple mini-tests. Students would not have to be shifted into different groups as all would take the same mini-test. Scoring would be simplified as all examinees would be given the same items.

The single item-sample approach would have the disadvantage of a somewhat greater measurement error than would be expected with the administration of a complete test or multiple matrix-samples. It would not be appropriate for some testing situations requiring greater accuracy.

The single item-sample approach could be useful in a situation where a teacher or school administrator wanted an estimate of the status of a group of pupils in a particular subject area but was reluctant to invest the time required to administer a complete standardized test. Instead of the more complicated multiple matrix-sampling approach, the use of a single mini-test would provide a simple, practical means for estimating group achievement with a minimum of time and effort.

In studies concerning item-sampling which have been published to date, the problem of "goodness" of an estimate of a population parameter has been treated in a variety of ways. Most studies have treated the differences between estimated and actual parameters in a descriptive fashion without attempting to establish criteria for judging the effectiveness of the sampling method utilized. More research is needed to provide information for the person considering the use of item-sampling to evaluate whether or not a particular method will provide suitable estimates of group performance.

II. BACKGROUND OF THE STUDY

Genesis of the Problem

The problem undertaken in this investigation had its origin in an evaluation project conducted by the Office of Teacher Education of The City University of New York (CUNY) at an experimental, public elementary school in the Bronx. As a part of the evaluation team, this investigator became aware of the extensive amount of classroom time used for the administration of standardized tests in public schools. This loss of teaching and learning time resulted in a disruption of curriculum plans with the pupils being the inevitable losers.

In addition to the Metropolitan Achievement Test (Form H) which was administered in reading and in mathematics, as required by the City of New York, pupils of the experimental school were also tested by two private educational firms which had been awarded evaluation contracts in conjunction with funds received by the school. Rather than carrying out yet a fourth standardized testing program, it was mutually agreed upon by the administrators of the school and the director of the CUNY project that other approaches to evaluation would be considered.

Teachers and administrative personnel at the school were interested in the development of tests which they hoped would be appropriate for use by their pupils and more compatible with their curricular goals than were existing standardized tests. A test development project was undertaken by school personnel with the assistance of the CUNY team. In addition, a literature search was conducted by this investigator of existing item-sampling methods used to estimate group performance on a complete test. It was found that the prevailing approach now in use was that of matrix-sampling, a logistically complex although apparently effective method of deriving estimates of group achievement.

The primary objective of this study was to determine the possibility of using a single sample of items drawn at random from a standardized test to provide estimations of a population mean and standard deviation on a total test. The use of a single item-sample, or mini-test, would provide a more practical approach to group evaluation in a school setting than the use of multiple matrix-sampling requiring several mini-tests taken by different samples of examinees.

The use of a mini-test instead of a complete test would not eliminate the objections often voiced against standardized tests. However, it would considerably reduce the amount of time and money spent on this type of testing, and should prove useful in the assessment of group performance in situations where reference to norms provided for standardized tests is of importance.

Review of the Literature

The literature concerning the use of item-sampling for the estimation of group achievement on a standardized test can generally be divided into two categories. The first category consists of post hoc sampling studies and the second category contains empirical studies in which item-samples are selected before they are administered. A majority of the studies to date come under the former classification.

In post hoc item-sampling investigations (Johnson and Lord, 1958; Lord, 1962; Plumlee, 1964; Cook and Stufflebeam, 1967; Husek and Sirotnik, 1968; Knapp, 1968; Shoemaker, 1970a, 1970b, 1971a, 1971b; Pugh, 1971; Barcikowski, 1972; and Kleinke, 1972), the first step was to obtain item responses to a complete test for a population. Parameter estimates were then generated using various matrix-sampling procedures on the data obtained from the administration of the whole test. These estimates were then compared with the actual parameters for the whole test. In a number of studies (Johnson and Lord, 1958; Lord, 1962; Plumlee, 1964; Cook and Stufflebeam, 1967; Husek and Sirotnik, 1968; Knapp, 1968; Pugh, 1971; and Barcikowski, 1972), matrix-sampling estimates were also compared with estimates of parameters derived through examinee-sampling, in which a random sample of examinees were administered a complete test to obtain estimates of the total population's performance on the test.

With the exception of two post hoc item-sampling studies (Johnson and Lord, 1958; and Cook and Stufflebeam, 1967) which did not utilize random sampling of both examinees and items, the item-sampling investigations reviewed here have taken the form of matrix-sampling.

The small number of empirical item-sampling studies reported in the literature (Owens and Stufflebeam, 1969; Cahen, Romberg and Zwirner, 1970; and Sirotnik, 1970) have used the procedure of obtaining population responses to a complete test and also to item-samples using a multiple matrix design. Both a complete test and item-samples were actually administered, in contrast to post hoc studies.

Post Hoc Studies

An early study by Johnson and Lord (1958) compared the efficiency of item-sampling with examinee-sampling in estimating a group mean. The data base for the study consisted of item scores for 160 college students on an 11 item calculus test, from which data for only 33 students were selected randomly for the purpose of their investigation. For the examinee-sampling, all 33 students were scored on each of the 11 items and a mean was calculated for each item. In the item-sampling procedure, the 11 items were randomly assigned to the 33 students so that 3 students responded to each of the items. This was repeated 30 times with different assignments of items to students. This differential assignment of items produced more consistent mean values and was said to be an improvement over the usual procedure of assigning the same item to each student. The authors concluded that this new procedure was valuable when testing time was limited and when group performance, rather than individual performance, was to be evaluated.

In a later study by Lord (1962), estimates of a population mean, variance and norms distribution were carried out using data for 1,000 examinees who had taken a 70 item test. Lord subdivided the population into 10 nonoverlapping groups of 100 examinees each. Item-samples of 7 items each were drawn with replacement and at random from the total 70 items. Comparisons of means, variance and distributions (using negative hypergeometric distributions) were made between item- and examinee-samples and the norms population. Values for the 10 item-samples were averaged to obtain a single estimate of the population parameters. The item-sample estimate was shown to be closer to the norms population value than any of the examinee-sample estimates.

Lord pointed out that in item-sampling an assumption is made that administering a sample of items out of context of the full test does not affect the examinee's performance too greatly. He suggested that this assumption is not justified in some practical situations.

Plumlee (1964) tested Lord's item-sampling technique for estimating means and standard deviations for smaller groups of examinees. A norms population of 200 adult examinees and 30 test items were used. Item-samples consisted of 10 random subgroups of 20 examinees for which data on 3 items selected randomly were used. Examinee-sampling procedures used the same 10 subgroups and the total test of 30 items. The item-sampling method proved to be more effective in estimating the mean, but generally less effective in estimating the standard deviation of the norms population than was the examinee-sampling procedure. However, the standard deviation estimated by item-sampling was only about one raw score point in error, which Plumlee stated was within the error of measurement expected if another form of the total test had been given. Plumlee suggested that item-sampling be utilized in developing test norms.

Cook and Stufflebeam (1967) investigated item-sampling procedures for estimating population means, standard deviations and score distributions using different sizes of item-samples and examinee-samples to compare the effectiveness of the two sampling methods. The data for this study were obtained from the administration of a 115 item hygiene test to 1,239 examinees. Estimates of norms were made using samples of 10, 25, 33 and 50 per cent of the total item and examinee populations. Item-samples were drawn without replacement from the pool of 115 items using a table of random numbers. It appears that item responses for the entire examinee population were utilized for each item-sample, rather than using the multiple matrix-sample design. Examinee-samples were selected by taking every tenth subject for the 10 per cent sample, every fourth subject for the

25 per cent sample, and so on for the 33 and 50 per cent samples. Estimation of the population distribution was carried out by fitting negative hypergeometric curves to the item- and examinee-samples.

Results from all four sizes of item-samples and examinee-samples provided "good" estimates of the population mean and variance. "Good" was defined as within one standard error of measurement of the population parameters. Nearly all of the hypergeometric distributions were good approximations of the norms distribution as determined by Chi-square values, with the item-sampling estimates generally superior to the examinee-sample estimates.

Kuder-Richardson 20 (KR20) and 21 (KR21) internal consistency coefficients were given for each of the examinee-samples and were essentially equivalent to the population value. Reliability coefficients were not given for the item-samples, nor was the problem of reliability in the use of item-samples discussed. As all items were taken by the examinee-samples and only portions of the total test were given in the item-samples, reliability would seem to be an even more important consideration in the latter procedure.

Husek and Sirotnik (1968) conducted a post hoc matrix-sampling and examinee-sampling study using the responses to both a 140 item final examination for 200 students (Test A), and a 30 item achievement test for 350 students (Test B). Matrix-samples for Test A consisted of 20 examinees and 14 items each; and, Test B had 35 examinees and 3 items each.

The examinee-samples combined the examinees from every two matrix samples. The matrix-sampling approach was found to be generally more efficient than the examinee-sampling method for estimating a population mean and variance. The authors suggested caution be employed when using matrix-sampling to estimate variance for tests designed to measure attainment of objectives, as such tests often minimize inter-subject variability.

Knapp (1968) applied balanced incomplete designs to the estimation of mean, variance and KR20 coefficients for a test. The data used were the first 29 items of a 100 item aptitude test administered to 4060 secondary school students. Ten random samples each of 3, 5, and 7 items were selected and the population was divided into 10 random samples of 406 students each. The balanced incomplete block design was utilized to give 10 estimates of population means, variance and reliability coefficients for each sample size. Examinee-sample estimates were also obtained using all 29 items for each sample of 406 students.

Results showed that the estimates of the population parameters were "quite good" for the item-samples, although the examinee-sample estimates of population values were better. It was concluded that 3 items per individual seemed to be sufficient and the additional accuracy acquired through examinee-sampling was not worth the greater amount of testing time involved.

Knapp (1968) stated that because each of the total test characteristics (mean, variance and reliability) is a function of item statistics only, it seemed permissible to estimate these characteristics based on the administration of item-samples. It would appear that an important consideration related to this assumption, and not mentioned by Knapp, is that the administration of item-samples rather than a total test may have an effect on the item statistics causing them to vary from what they would be for a total test, thus affecting the parameter estimates.

Shoemaker (1970a) conducted an investigation of the estimation of a norm distribution by post hoc item-sampling varying the number of subtests, number of items per subtest and the number of examinees responding to each subtest. The norm distribution consisted of test scores for 810 college students on a 150 item multiple-choice psychology test. Results showed that the variable of importance in estimating a norm distribution (using negative hypergeometric distributions), mean and variance was not the item-sampling procedure, but the number of observations obtained. An observation was defined as the score received by one examinee on one item. The results showed that as the number of observations increased beyond 1.23 per cent of the data base (the number of observations in the norm distribution) all procedures produced similar results.

Using estimates of the parameters obtained by the item-samples, the KR21 reliability coefficient for the full-length test was computed for each sampling procedure. The estimated KR21 coefficients were very close to the actual value for the total test, and in many cases were higher. Again no discussion was included of the actual reliability of the item-samples.

In another study using a matrix-sampling design, Shoemaker (1970b) dealt with the problem of estimating a negatively skewed and a normally distributed score distribution. One set of norm data consisted of test scores with a markedly negatively skewed distribution for 1031 first graders on a 20 item multiple-choice test. The item-examinee samples were systematically varied in number of subtests, number of items per subtest and number of examinees per subtest. Each sampling procedure was replicated five times.

A normal distribution of scores for 1031 examinees on a 20 item test was generated using a Monte Carlo approach. The same item-examinee sampling procedures were carried out as with the data for the skewed distribution.

Shoemaker again found that the degree of accuracy in predicting a population mean and standard deviation is a function of the number of observations. Results were not significantly influenced by the degree of skewness in the normative distribution.

An estimated KR21 coefficient for each sampling procedure was computed using the estimated means and standard deviations derived from the matrix-sampling. The estimated reliability coefficients for the whole tests were very close to the actual total test values.

In some of the sampling procedures utilized by Shoemaker, a number of items were excluded from all subtests including three samples where only 10 of the total 20 items were administered. Although Lord and Novick (1968, p. 257) have indicated that failure to administer all test items increases the standard error of estimating the population mean, no trend was apparent in Shoemaker's study (1970b) to support this statement when standard errors were examined as a function of the number of items omitted.

Shoemaker recommended that in all sampling investigations standard errors of estimate for each parameter be included to aid in the interpretation of results.

In another study, Shoemaker (1971a) investigated the standard error in estimating a population mean and variance with matrix-sampling procedures having the same number of observations for a given test length. A normal and a very negatively skewed normative distribution of scores were used as a data base. The scores for the normal distribution were computer-generated using a Monte Carlo procedure.

Three levels of number of observations, 1500, 3000 and 4500, were selected and three matrix-sampling procedures were replicated 300 times.

Results showed that each procedure having the same number of observations had, in general, the same standard error in estimating variance. This was not true when estimating a population mean. A significant factor in accounting for differences in the standard errors of estimating the mean was the variance of the item difficulty indices. It was recommended that a large number of subtests be used and that a few examinees be administered many items rather than many examinees taking few items for some values of item difficulty variance.

Shoemaker (1971b) also investigated the use of matrix-sampling for scaling attitudes by the method of paired-comparisons. The data base consisted of responses by 407 primary grade pupils to a 15 item test designed to scale degree of affect to six stimuli. The degree of similarity between the sample scale values and normative scale values was quantified in terms of the generalized distance function "D". For any two sets of scale values, "D" was defined as the square root of the sum of the squared differences between scale values for each stimulus. Results showed that matrix-sampling produced satisfactory estimates of scale values. Estimation was shown to increase in efficiency as the sample size, both in terms of number of items and examinees, and the number of samples increased.

A study reported by Pugh (1971) compared the effectiveness of matrix-sampling and examinee-sampling in estimating means and standard deviations for an attitude scale consisting of Likert items. The data for the study consisted of the responses of 600 fifth and sixth grade pupils to a 60 item scale in the area of student satisfaction with school. Sample sizes of 5, 10 and 20 per cent were used for both the matrix-sampling and examinee-sampling. Matrix-sampling was found to be generally more efficient than examinee-sampling, and larger samples more efficient than smaller samples in the estimation of means and standard deviations.

A Monte Carlo approach was used by Barcikowski (1972) to generate item responses for different sets of item parameters,(range of difficulty and range of biserial correlation). The efficiency of estimates of means and variances of distributions arrived at by matrix-sampling were compared with estimates from examinee-sampling. Item-sampling proved superior under all conditions for estimating means. Examinee-sampling provided better variance estimates than item-sampling for certain test lengths, ranges of item difficulty and discrimination.

In the Barcikowski (1972) study, estimates ($\hat{\theta}$,s) of population mean and variance values ($\hat{\theta}$'s) were made for each sampling plan and a sum of squared errors (SSE) was computed: $(\hat{\theta}-\theta)^2$. The ratios of the SSE between item and examinee samples were then examined to see which was more efficient as a method of estimation.

Kleinke (1972) has investigated a method for approximating total-test distributions through matrix-sampling that is an alternative to the negative hypergeometric approach; it is the linear-prediction (LP) approach. With this method, item-sample responses are used to predict the score of an individual on the remainder of the total-test. The sum of this predicted score and the obtained item-sample score is the predicted total-test score. The LP distribution is the distribution of the predicted total-test scores for all examinees. The LP approach proved effective in this study, although its generalizability was not established.

Empirical Studies

Sax and Cromack (1966) conducted an empirical study of the effects on test performance of different arrangements of items according to difficulty. Although this was not a study utilizing item-sampling to estimate group performance, it was of interest as it dealt with item arrangement, one aspect of context effects, an important consideration in item-sampling.

The subjects included 276 students in a college education course and 191 students in a college psychology class. They were given test forms with items arranged from easy to hard, hard to easy, hard regularly interspersed with easy, and random order of difficulty. The results generally supported the hypothesis that items should be arranged in order of ascending difficulty if tests were long or time limits restricted. If time limits were generous, there was no advantage in arranging items by difficulty.

Owens and Stufflebeam (1969) compared the accuracy of item-sampling and examinee-sampling for estimating test means and standard deviations when samples were drawn prior to test administration. Item-samples consisted of 3, 6 or 12 items randomly selected from the 50 multiple-choice items of the Metropolitan Elementary Reading Test, Word Knowledge subtest. The study was replicated on two school populations, one advantaged and one disadvantaged. Item-samples were administered to 376 elementary school students in each of three subgroups. Examinee-sampling consisted of random samples of 6, 12 and 24 per cent of the total examinee group of 539. Another group took the entire test to provide one criterion for judging the estimates of test norms. After completing the item-samples, the pupils in the item-sampling groups then answered the remainder of the items in the test which were included in their test booklets. The other criterion for judging the accuracy of the estimates consisted of the norms developed from scoring all items for all pupils in the study.

Differences between estimated means and standard deviations and actual population means and standard deviations were described in terms of standard deviation units. The accuracy of estimates of the mean were evaluated in terms of the standard errors of the sampling distribution of the means calculated using formula 7.17.2 from Hays (1965). For lower ability disadvantaged students, all mean estimates were within one standard error of the criterion means. For the higher ability advantaged group, some mean estimates were outside two standard errors of both criterion means.

The accuracy of the estimations of the standard deviations was checked using formula 11.6.1 from Hays (1965) for testing exact hypotheses about a single population variance. None of the estimated standard deviations exceeded the 95 per cent confidence limits for either criterion.

Estimates of group means and standard deviations were as good for the item-samples as for those obtained from examinee-samples. Results indicated that both sampling methods may operate differently for higher and lower ability students, with less precision in estimates of the mean for higher ability students from advantaged neighborhoods. The findings also indicated that the variation of item sequence does not necessarily have a significant effect on test performance.

Cahen, et al (1970) reported a study utilizing matrix-sampling in which each examinee was given a sample of items from a total test in addition to taking the complete test. The purpose of the item-sampling was to estimate school means for ninth grade mathematics students. Forms A and B of the 50-item Cooperative Arithmetic Test were administered at 81 schools. The two forms were nominally parallel. Form B was item-sampled on the first day of testing. The item-sample test booklets contained 9 items, 5 from the Cooperative Arithmetic Test and 4 from two other arithmetic tests. The data gathered for the other arithmetic tests were to be used in another study. Form A was administered in its usual complete version on the second day of testing. Time limits were strictly observed for the administration of both the total test and the item samples, although pupils were given a slightly longer time proportionally, by 30 seconds, on the item-samples than the complete test.

Results showed the correlation between item-sample estimates and means obtained from administration of the total test for the schools exceeded .80. Item sample means for Form B were consistently higher than the total score means for Form A. One explanation offered for this bias in favor of the short set of items was that it may have been due to motivational variables.

Cahen, et al concluded that the item-sampling technique is promising for the estimation of total group performance when the purpose of the sample administration is to estimate the relative rather than the absolute difference between schools.

In a study by Shoemaker (1970c), the problem of context effects was considered, although item-sampling was not a part of the experimental design. Shoemaker examined the effect on mean test score, item difficulty index and reliability and validity coefficients of reordering items within a test. A 10 item letter-series-completion power test and a 20 item number-series speed test were used. The subjects were 139 college students. Reordering effects were found to be generally minimal. Shoemaker concluded that these findings implied that the context effect on item-sampling performance was also minimal.

This study did not consider factors involved in context effects which are related to the length of a test such as differential motivation, fatigue and practice effects.

These factors must also be considered before drawing conclusions about the presence or absence of context effects in the use of item-sampling.

A study by Sirotnik (1970) compared estimates of group means and variance derived by post hoc matrix sampling and by actual a priori matrix sampling. The population consisted of 180 college students. The tests used were 30 item vocabulary, mathematics and attitude tests. The students were randomly divided into two groups. One group took the 30 item vocabulary test on the first day (Treatment A), and the other group took one of ten different item-samples consisting of 3 items from each of the tests (Treatment B). The students taking the item-sample then took the entire 30 item vocabulary test. On the second day all of the subjects took the entire mathematics test, and on the third day all took the 30 item attitude test. Strict time limits were enforced although students were judged to have ample time to complete the whole tests and the item-samples.

Examinees receiving Treatment A were scored on only one of the ten 3 item samples although they took the entire test. Examinees in Treatment B took the item-sample packet itself. Scores from both Treatments were used to provide parameter estimates. An analysis of variance showed no significant difference in estimates of mean and variance between Treatments A and B for the three tests.

The associated 95 per cent confidence intervals for the actual mean differences for Treatment groups A and B were given to aid in the interpretation of mean differences although Sirotnik suggested that caution be exercised in interpreting these differences. He concluded that matrix-sampling produced mean estimates relatively immune to context effects and that variance estimation should be studied further.

It is possible that some psychological factors involved in context effects were not present because subjects in Treatment B took the item-samples followed by a complete test, rather than simply taking an item-sample as a substitute for a longer test.

If confidence intervals are to be used to evaluate mean differences between treatments, then confidence intervals for the estimated means should be included as well as for actual means. Both the actual parameter and the estimate of that parameter include some amount of measurement error.

Summary of Literature Review

The studies reviewed here have been concerned primarily with the investigation of item-sampling techniques through poct hoc sampling of items from total-test data. A number of the investigations have compared item-sampling with examinee-sampling, and most have shown item-sampling to be either equally as effective as examinee-sampling or more effective. A summary of the more pertinent details of the studies reviewed is presented in Table 1.

With the exception of one study by Shoemaker (1970b), investigation of the effect of the total omission of items in an item-sampling procedure has not been reported. The results of Shoemaker's study indicated that the use of item-samples which omit some items from a total test does not seem to affect estimates of group performance. The use of a single sample of items for estimating group performance, with the omission of the remainder of the items in the test, is the subject of this dissertation project.

Much of the literature concerning the use of item-sampling gave no consideration to the problem of the reliability of the sampling estimates. Those few studies which included the use of reliability coefficients gave the estimated reliability of the total test based upon the item-sample estimates. The reliability of the item-samples themselves was not mentioned in any study. The reliability of the samples of test items may be used in evaluating the estimates of population parameters derived from the use of a sampling procedure.

Three empirical studies (Owens and Stufflebeam, 1969; Cahen et al, 1970; and Sirotnik, 1970) were designed to investigate possible context effects on performance in the use of item-samples. The Owens and Stufflebeam study and

TABLE 1

Summary of Item-Sampling Studies Reviewed

Author and Date	No. of Subjects: Total	No. of Subjects: Per Sample	No. of Items: Total	No. of Items: Per Sample	Population	Estimated Reliability[a]	$\hat{\mu}$ & $\hat{\sigma}$ Goodness Criterion
			Post hoc studies				
Johnson & Lord, 1958	33	3	11	1	college	no	Descriptive & ranking
Lord, 1962	1000	100	70	7	college	no	Discrepancy index D[b]
Plumlee, 1964	200	20	30	3	adults	no	Descriptive & error of meas. for par. form
Cook & Stufflebeam, 1967	1239	124, 310, 409 & 620	115	12, 29, 38 & 58	college	no	Within one SEM & Chi-square[b]
Husek & Sirotnik, 1968	200 350	20 35	140 30	14 3	college college	no no	Descriptive
Knapp, 1968	4060	406	29	3, 5 & 7	secondary school	yes	Descriptive

[a]Estimated reliability for total test presented in study.

[b]Criterion for estimated normative distribution.

TABLE 1 - Continued

Author and Date	No. of Subjects: Total	No. of Subjects: Per Sample	No. of Items: Total	No. of Items: Per Sample	Population	Estimated Reliability[a]	$\hat{\mu}$ & $\hat{\sigma}$ Goodness Criterion
			Post hoc studies-continued				
Shoemaker, 1970a	810	10 & 20	150	15, 30, 50 & 75	college	yes	Descriptive & Kolmogorov -Smirnov[b]
Shoemaker, 1970b	1031	30, 60, 90 & 120	20	5, 10 & 15	Monte Carlo	yes	S.E.-mean
	1031	30, 60, 90 & 120	20	5, 10 & 15	first grade	yes	S.E.-st. dev.
Shoemaker, 1971a	?	30, 60, 120 & 180	?	5, 10 & 15	Monte Carlo	no	S.E.-mean &
	?	30, 60, 120 & 180	?	5, 10 & 15	?		S.E.-st. dev.
Shoemaker, 1971b	407	15 & 30	15	3, 5 & 10	first grade	no	Gen. distance function D[b]
Pugh, 1971	600	30, 60, & 120	60	3, 6, & 12	fifth & sixth grades	no	Descriptive
Barcikowski, 1972	10010	385	97, 150 & 194	5 & 6	Monte Carlo	no	Sum of squared errors
Kleinke, 1972	167000	105	100	10	twelfth grade	no	Chi-square[b]

[a]Estimated reliability for total test presented in study.
[b]Criterion for estimated normative distribution.

TABLE 1 - Continued

Author and Date	No. of Subjects: Total	No. of Subjects: Per Sample	No. of Items: Total	No. of Items: Per Sample	Population	Estimated Reliability[a]	$\hat{\mu}$ & $\hat{\sigma}$ Goodness Criterion
Empirical studies (context effects investigated)							
Owens & Stufflebeam, 1969	1903 2051	22, 47 & 94 22, 47 & 94	50 50	3, 6 & 12 3, 6 & 12	fourth grade	no no	Formulas 7.17.2 & 11.6.1, (Hays, 1965) S.E.-mean 95% conf. limit
Cahen et al, 1970	81 schools	10 to 505	50	5	ninth grade	no	Descriptive
Sirotnik, 1970	180	9	30	3	college	no	ANOVA & 95% conf. interval
Studies of item arrangement							
Sax & Cromack, 1966	276 191	69 48	70 70	70 70	college college	no no	ANOVA, ANOCOVA & descriptive
Shoemaker, 1970c	139 139	23 to 40 23 to 40	10 20	10 20	college college	yes	ANOVA & descriptive

[a] Estimated reliability for total test presented in study.

the investigation by Sirotnik led to the conclusion that the context in which items were presented did not affect performance on the items. However, in both studies, the item-samples were administered as the initial portion of a longer test, thus possibly confounding the evaluation of psychological factors in context effects present in the use of isolated item-samples. In the study by Cahen et al, (item-samples) were administered separately from the remainder of the test. The findings of the study showed that the estimated means for the item-samples were consistently higher than the means for the total test.

The Owens and Stufflebeam (1969) study used a population of elementary school pupils, Cahen et al (1970) used ninth grade pupils and Sirotnik (1970) used college students. Further investigation of the possible context effects involved in the use of item-sampling for different populations is needed.

The problem of judging the effectiveness of sampling procedures has been dealt with in a variety of ways in the studies reviewed here, and has been omitted entirely in a number of reports. In some investigations the estimates of group parameters have been compared with actual population parameters in descriptive terms, such as being "fairly close", "good estimates", or within a number of standard deviation units. Other studies have set criteria for sample estimates in terms of standard errors of measurement for a test, standard errors of a parameter, or confidence limits for the estimation of a parameter. Shoemaker (1971b) has approached the problem of evaluating the effectiveness of a sampling procedure by generating a sampling distribution of estimated parameters in order to arrive at the standard errors associated with a particular sampling method. This approach has been adopted in evaluating the use of a single sample of items in the study reported here.

Statement of the Problem

The use of item-samples to estimate group performance is one solution to the problem of the time involved in the administration of standardized tests. A review of the literature has revealed a number of problems associated with item-sampling. The following problems have been selected for study in this investigation:

1. Can a single sample of items selected at random from a standardized test provide an efficient estimate of a group mean and standard deviation?
2. Does a comparison of post hoc sampling estimates with empirical (a priori) sampling estimates, using a single item-sample, result in data supporting the common assumption that context effects on performance in the use of item-sampling do not exist?
3. How effective is the use of a single item-sample as determined through post hoc sampling?

4. How reliable are the actual item-sample statistics, as well as the estimated parameters derived from the item-sample data?

III. METHOD

This study consisted of two phases. In the first phase, the effectiveness of different sizes of random samples of items drawn from several standardized tests was investigated through post hoc item-sampling. In the second phase, the item-sampling procedure which was most efficient in the first phase of this study was used for empirical (a priori) and post hoc item-sampling investigation. The use of a single item-sample for estimating group performance on a total test was studied in contrast to the more commonly used approach of matrix-sampling.

The initial phase of this study consisted of post hoc analyses of item samples selected from five standardized reading and mathematics tests for which item data were available from The City University of New York Open Admissions Pilot Project. Items were selected from the tests by a random sampling procedure.

Item-sampling studies reported in the literature have used sample sizes ranging from 6 to 50 per cent of the total test (Cook and Stufflebeam, 1967 and Owens and Stufflebeam, 1969). For this study, samples of 5, 10, 25 and 33 per cent of the total number of items in each test were selected. Samples greater than 33 per cent were not used as they would minimize the practical benefits gained by administering a sample of items rather than an entire test.

The tests used and their population sizes were: 1) Davis Reading, Form 1A, 105 subjects; 2) Davis Reading, Form 2A, 94 subjects; 3) STEP Reading, Form 2A, 100 subjects; 4) STEP Math, Form 2A, 90 subjects; and, 5) Stanford Achievement Math: Computation, Form W, 196 subjects.

Means, standard deviations and standard errors of measurement were determined for each test. For each item sample, an estimated mean and standard deviation for the whole test were calculated using extrapolation formulas developed by Owens and Stufflebeam (1969), from Lord (1968):

$$\hat{m} = \frac{R}{r}(\bar{x})$$

$$\hat{\sigma} = \sqrt{\frac{n}{r(r-1)(n-1)}\left[R(R-1)\left[\frac{\Sigma x^2}{n} - \bar{x}^2\right] - R(R-r)\left[\bar{x} - \Sigma\pi^2\right]\right]}$$

$\hat{m}$ - estimated mean for full length test

R - total items in entire test

r - number of items in each sample

$\bar{x}$ - composite mean for item samples (in this study, the single sample mean)

$\hat{\sigma}$ - estimated standard deviation for full length test

n - number of subjects taking each R/r item-sample

x - raw score for an individual

π - proportion of n subjects answering an item correctly

The formulas developed by Owens and Stufflebeam to extrapolate item-sample means and standard deviations to obtain estimates of population parameters were used in a matrix-sampling study. They are comparable to formulas presented by Shoemaker (1971c) for use with single item-samples which were then pooled in matrix-sampling to provide average estimates of parameters. In this study, the formulas were used for single-sample estimates only, rather than in a matrix-sampling context.

For each of the five tests, comparisons were made between the estimates of the population means and standard deviations for each item-sample and the actual population parameters for the CUNY Open Admissions data. This comparison of estimated with actual parameters was carried out to determine the smallest item-sample size which would consistently provide a "good estimate". The criterion for a "good estimate" of a population mean was defined as one that was within plus or minus one standard error of measurement (SEM) of the actual population parameter (Cook and Stufflebeam, 1967).

The means, standard deviations and standard errors of measurement, as well as the estimated means and standard deviations for the five tests are presented in Table 2. The smallest random sample size that met the good estimate criterion for each of the five tests was the 25 per cent sample. All five estimated means were within plus or minus one SEM of the population means. The estimated standard deviations also compared favorably with the population standard deviations with less than one raw score point difference between estimations and actual parameters. The 33 per cent samples also met the criterion for each of the five tests. The 10 per cent sample size gave a good estimate of the mean for four of the tests, and the 5 per cent sample was satisfactory for three of the tests. Thus, the 25 per cent sample size was the smallest random sample which consistently met the criterion for a good estimation of the mean for the five standardized tests used in the CUNY Open Admissions project.

Empirical Study - Arithmetic Test

The second phase of this study was designed to test the effectiveness of using a single 25 per cent random sample of items, in an a priori situation, to estimate a group mean and standard deviation on a whole test administered in a classroom setting. Special emphasis was placed upon investigation of the effect of the context in which an item sample was administered on pupil performance. The estimation of pupil performance on a test was carried out with an item sample from the same test as well as with an item sample from a parallel form of the test.

Subjects

The subjects were 204 second grade pupils from an elementary school in the Bronx. The pupils were from eight classes, and had been assigned to their classes at the beginning of the school year according to their records of achievement so that each class would be homogeneous in ability. For the purpose of this study, the classes were divided into four groups, with two classes per group. An attempt was made to see that there would be approximately

TABLE 2

Estimation of Group Parameters with Random Item-Samples of Different Lengths

Test[a]	No. of Subj.	No. of Items	S.E.M.	Total Test		Random Samples							
						5%		10%		25%		33%	
				m	σ	$\hat{m}$	$\hat{\sigma}$	$\hat{m}$	$\hat{\sigma}$	$\hat{m}$	$\hat{\sigma}$	$\hat{m}$	$\hat{\sigma}$
1	105	40	2.73	19.20	8.14	23.20[b]	2.07	20.31	9.62	20.00	8.35	19.69	8.74
2	94	40	2.51	28.93	6.84	27.40	7.83	25.60[b]	6.12	29.16	7.65	30.86	6.53
3	100	70	3.52	36.29	16.39	27.83[b]	16.25	37.40	16.66	38.38	17.15	39.32	16.65
4	90	50	3.11	24.86	8.83	24.33	8.41	27.00	8.03	24.69	7.93	25.06	8.41
5	196	41	2.58	24.82	10.15	23.99	11.55	24.91	9.59	23.29	10.48	24.92	10.18

[a]Test 1 - Davis Reading (Form 1A)

Test 2 - Davis Reading (Form 2A)

Test 3 - STEP Reading (Form 2A)

Test 4 - STEP Math (Form 2A)

Test 5 - Stanford Advanced Math: Computation (Form W)

[b]Estimated mean did not fall within plus or minus one SEM of population mean.

the same average of high ability and low ability pupils in each group according to the assistant principal's ranking of the classes. It was recognized, however, that the assignment of pupils to classes at the beginning of the school year according to some measure or measures of achievement would very likely be invalid for a number of pupils by the end of the year. The groups of pupils consisted of: Group A, 48 pupils; Group B, 53 pupils; Group C, 55 pupils; and, Group D, 48 pupils.

Procedures

The Stanford Achievement Test, Test 6: Arithmetic (Stanford Arithmetic Test), Primary I, Form X, (Harcourt, Brace Jovanovich, Inc., 1964-1966) was administered to all of the subjects by their own classroom teachers. Standardized administration instructions were followed, with one exception. Teachers were told that the time limits were approximate and that most of their class should be given the opportunity to complete the test. The reason for this exception to standardized procedures was that the administration of item samples to estimate performance on a speed test is not as valid as for a power test (Lord 1962, 1968). All tests which were not completed were omitted from the data analysis (Plumlee, 1964).

Two mini-tests were constructed from the total Stanford Arithmetic Test of 63 items. One was selected from Form X and one from a parallel form, Form Y. The complete test consisted of four sections with different types of items in each section. A random sample of 25 per cent of the items was selected from each section. The mini-test from Form X (mini-test, Form X) and the mini-test from Form Y (mini-test, Form Y) each consisted of 16 items. The purpose of constructing a mini-test from Form Y as well as from Form X was to determine which mini- test would give a better estimate of student performance on the total test of Form X in this experimental situation. It might be hypothesized that performance on mini-test, Form X would be better than on mini-test, Form Y as the pupils were familiar with the items from having taken the total test, Form X, previously. Thus, it might seem more reasonable that students be given items from a parallel form with which they were not acquainted. However, the problem remained that tests which are nominally parallel in the sense of classical test theory, that is tests whose means, standard deviations and correlations with any and all outside criteria are equal (Gulliksen, 1950), are in reality non-existent. Tests that are not identical cannot be absolutely parallel. To the extent that items differ, they measure different properties. Therefore, it was decided to use mini-tests from both Form X and from Form Y to determine which would provide the better estimate of group achievement on the whole test, Form X. In the study by Cahen, et al (1970) which employed traditional matrix sampling, item samples from a nominally parallel test Form B were utilized to estimate group achievement on Form A. The studies by Owens and Stufflebeam (1969) and by Sirotnik (1970) used item samples from the same test form as were administered *in toto*.

A small group of interested teachers and one teacher-trainer from a neighboring elementary school in the Bronx formed a committee to develop achievement tests which they hoped would be more appropriate for their pupil population than the currently available standardized tests. They were assisted in their project by this investigator, primarily in the form of guidance and of provision of technical services. This committee developed an arithmetic test and a reading test for the second grade level.

The teacher-made arithmetic test was used in conjunction with the mini-tests for two of the groups of pupils, C and D. Mini-tests were administered as a part of the longer teacher-made test for these two groups to investigate the possible context effect on pupil performance when item-samples are part of a longer, more complex test than the standardized test, as opposed to being administered alone. In both situations, the items were administered out of their original context, which was as a part of the Stanford Arithmetic Test.

Two weeks after the complete Stanford Arithmetic Test, From X was administered, the eight teachers administered the mini-tests to their classes. The two classes in Group A were given only mini-test, Form X. The two classes comprising Group B were given mini-test, Form Y. The pupils in the two Group C classes were administered mini-test, Form X as a part of the longer teacher-made arithmetic test. The two classes in Group D were given mini-test, Form Y as a part of the same teacher-made test. The two mini-tests and the teacher-made test are presented in Appendices I to III. The design for the mini-test administration is illustrated in Figure 1, along with the post hoc analysis described below.

Test administration instructions given to the teachers who administered the mini-tests are included in Appendices IV to VII. The teachers were again told that time limits given in the instructions were approximate guidelines and that most of the pupils should be given an opportunity to complete the test. Instructions for the administration of the mini-tests were adapted from the Stanford Achievement Test Directions for Administering the Primary I Arithmetic Test (Harcourt Brace Jovanovich, Inc., 1964-1966). Groups C and D were given their mini-tests as Test II of the teacher-made test. The approximate time given for taking the whole Stanford Arithmetic Test was 30 minutes, with one rest period. The approximate time for the teacher-made test with the mini-test was 38 minutes with one rest period after the third part of the test. Time guidelines for the teacher-made test were determined during pilot testing of the items with another group of pupils. The test consisted of the following parts: Test I (about 10 minutes), Test II - mini-test (about 9 minutes), Test III (about 4 minutes), Test IV

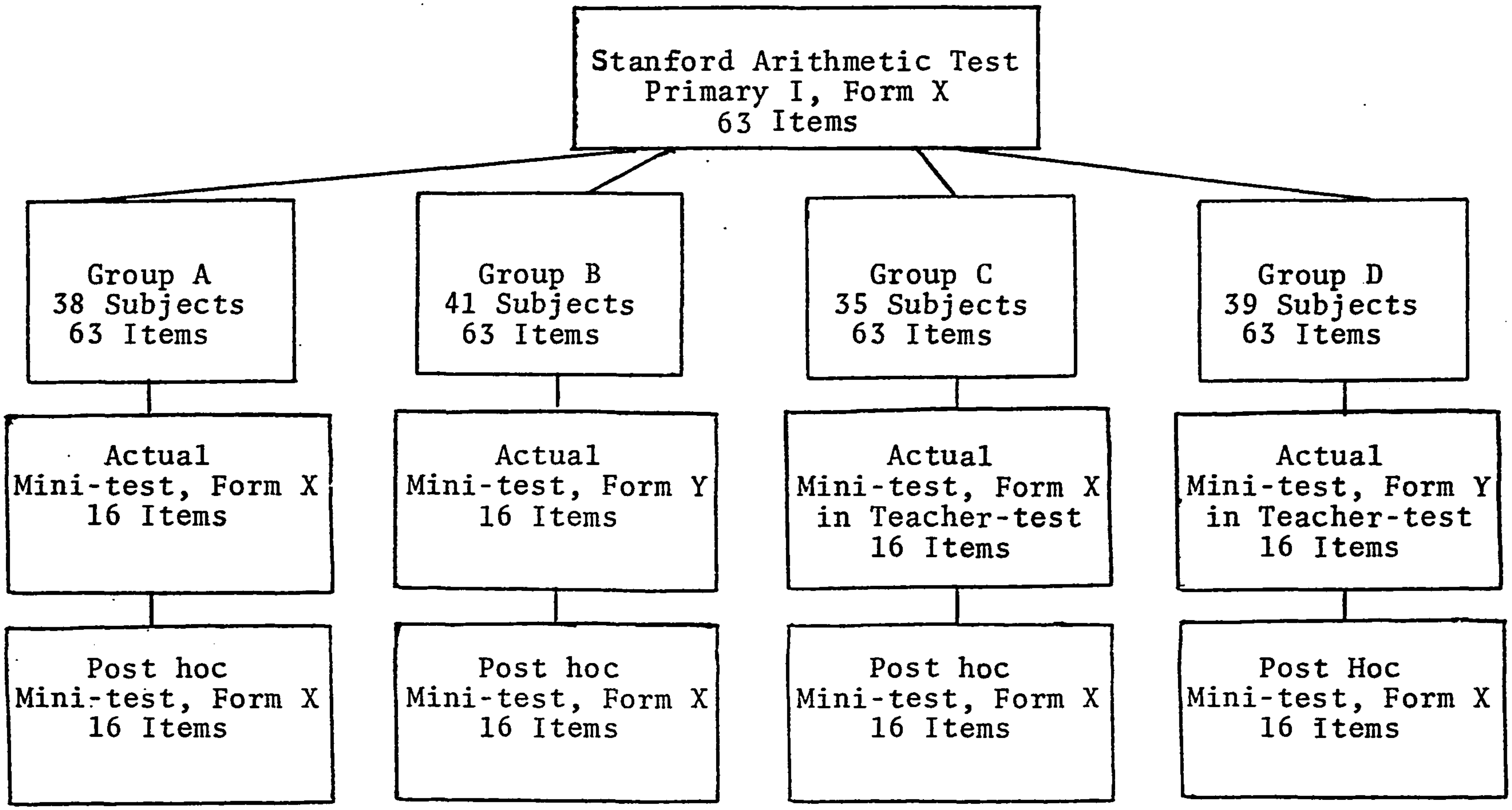


Figure 1. Design of Stanford Arithmetic Test Mini-test Administration.

(about 5 minutes), and Test V (about 10 minutes). Groups A and B, taking only the mini-test, were given approximately 9 minutes for the test, or longer if needed. The content of the teacher-made test included the following types of items: Test I, arithmetical operations using whole numbers and fractions; Test II, mini-test; Test III, patterns in number sequences; Test IV, questions presented orally; and, Test V verbal problem solving.

Mini-tests that were not completed were omitted from the data analysis, as was the case with the incomplete Stanford Arithmetic Tests. A number of pupils who had taken the complete test were absent when the mini-tests were administered. Tests for these pupils were also dropped from the final data analysis. As a result of the incomplete tests and absenteeism, the final numbers of subjects in each test group were: Group A, 38 pupils; Group B, 41 pupils; Group C, 35 pupils, and Group D, 39 pupils.

Post Hoc Sampling

The Stanford Arithmetic Test data were subjected to several post hoc item sampling analyses. For each of the four groups of subjects, a post hoc estimation of the group mean and standard deviation on the complete tests was obtained using an item sample containing the same 16 items included in mini-test, Form X. The pupils' performance on the whole test was compared with the estimates derived from this post hoc item sample and the estimates from the mini-tests.

This analysis provided information on the effect on pupil performance of the context in which items are administered. It also made possible the comparison of an item sampling procedure in both an a priori and a post hoc situation.

Further post hoc item sampling was carried out on the Stanford test data for the group as a whole, 153 pupils. Estimated means and standard deviations for the group's performance on the whole test were determined for each of the post hoc item samples drawn and compared with the actual mean and standard deviation for this population. The first item sample analyzed contained the same items included in mini-test, Form X. It resulted in an estimated mean well within the designated criterion and a "good estimate" as well of the standard deviation.

Ten more 16 item random samples were drawn in the same manner as the mini-test items, 25 per cent from each of the four sections of the Stanford Arithmetic Test. The purpose for analyzing the ten additional samples was to determine if the close estimation of the group parameters using the post hoc mini-test sample occurred by chance or if it could be duplicated consistently by similarly selected item samples.

Another ten 16 item random samples were selected, but not by the same procedure as the previous ten samples. Instead, the items, comprising 25 per cent of the total test, were selected at random from the whole test without regard to the four sections of the test. This procedure was carried out in order to determine whether it was necessary to select proportionately from the different parts of a test for a 25 per cent random sample of items to provide a good estimate of group achievement. The estimated means and standard deviations for each item sample were compared with the population parameters.

It was decided that the best description of the effectiveness of the item-sampling approach used in this study would be obtained by further post hoc sampling. In addition to the 10 samples drawn previously, ninety 25 per cent random samples, taken proportionately from each section of the test, were drawn. The estimated mean, standard deviation and KR21 reliability coefficient for each sample were determined. The formula used for estimating reliability was given by Shoemaker (1971c) for estimating this parameter from a single subtest:

$$\hat{KR21}_i = \frac{K}{K - 1}\left[1 - \frac{\hat{\mu}_i - \frac{\hat{\mu}_i^2}{K}}{\hat{\sigma}_i^2}\right]$$

K = the total number of items in the population.

$\hat{\sigma}_i$ = the estimated variance from subtest i.

$\hat{\mu}_i$ = the estimated mean from subtest i.

The frequency distributions for the three estimated parameters were plotted for the 100 samples and the standard deviations, or standard errors of estimate, were determined for each parameter.

Empirical Study - Reading Test

The administration of item samples from a reading test was also investigated for the purpose of determining whether the context in which a sample of reading test items were administered would have an effect on pupil performance. The effectiveness of the use of a single item-sample from a standardized reading test to derive an estimate of group achievement was also studied. The possibility that an item-sampling method may be test-specific was a consideration in using a reading test in addition to the arithmetic test described above.

Subjects

The subjects for this phase of the study were second grade pupils from two elementary schools in the Bronx. All second grade classes from one of the schools took part in this phase of the study, and some of the classes from the other school also volunteered to paricipate in this investigation.

Procedures

As a part of a city-wide testing program, the Metropolitan Achievement Test (MAT) in Reading (Harcourt Brace Jovanovich, Inc., 1970) was administered to all second grade pupils in New York City public schools. Approximately one month later, mini-tests derived from the MAT Reading Test were administered to pupils taking part in this study.

The Primary I level of the MAT was administered to some of the pupils and some were given the Primary II level, depending upon which test their teacher felt was more appropriate for them. The mini-test which was administered to them was drawn from the MAT Reading Test corresponding to the level which they had taken in its complete form. The pupils took Form H of the complete MAT Reading Test, and the mini-test items were drawn from the nominally parallel Form F.

Both the mini-test from the Primary I level of the MAT Reading Test (mini-test I) and the mini-test from the Primary II level (mini-test II) consisted of a 25 per cent random sample of items from the complete test. The Primary I MAT consisted of 42 items, the Primary II test had 44 items and each of the mini-tests had 11 items. The mini-test items were drawn proportionately from the three types of items in the MAT tests, one section consisting of sentences with two sections of stories.

A portion of the pupils taking mini-test I and mini-test II took the mini-test alone while some of the pupils took the mini-tests as part of a longer teacher-made test. The teacher-made reading test was constructed by the test committee from C.S. 234 described earlier. Mini-test I was administered to 113 pupils (Group I) and mini-test II to 26 pupils (Group II). Sixty-nine pupils (Group III) took mini-test I within the teacher-made test and 69 pupils

(Group IV) were given mini-test II with the teacher-made test. The teacher-made test was the same for both groups. The two mini-tests and the teacher-made test are included in Appendices VIII-X.

The reading mini-tests were administered to the pupils by their own teachers. Administration instructions for the mini-tests were adapted from the MAT Teacher's Directions (Harcourt Brace Jovanovich, Inc., 1970). The instructions for both mini-tests and the teacher-made test are given in Appendices XI to XIV. Teachers were instructed that the time limits given were approximate guidelines and that most of the pupils should have time to complete the mini-test. The assistant principals in charge of the second grades at both schools indicated that when the complete MAT Reading Test was administered to their pupils, it was their "unofficial policy" to give pupils ample time to complete the test rather than to adhere to the standardized test administration time limits.

The time period suggested in the Primary II Teacher's Directions for the MAT Reading Test is 30 minutes, and for the Primary I test is 30 minutes. The time guidelines for the mini-tests were proportionate to those given for the complete test. The suggested time given for the teacher-made portions of the test was determined by a tryout of the test items with other second grade pupils. The teacher-made tests with mini-tests consisted of the following parts: Test I (about 8 minutes), Test II (about 5 minutes), Test III - mini-test I (about 10 minutes) or mini-test II (about

8 minutes), Test IV (about 15 minutes), and Test V (about 10 minutes). The entire test took approximately 48 minutes for level I and 46 minutes for level II, with a rest period given after the third part of the test.

The first part of the teacher-made test consisted of aural discrimination items, the second part was made up of vocabulary items, the third part was the mini-test, stories made up the fourth part of the test, and the fifth part consisted of items with words that rhyme.

A small number of mini-test results had to be omitted because the tests were not completed. Test results for a few other pupils were also omitted from data analysis when it was learned that they had not taken the complete MAT Reading Test. The data for one class which was a part of Group III could not be used as administration instructions were not followed for that class. Parts I and II of the test were administered on one day with Part III, the mini-test, administered alone on another day for that class.

The final number of pupils in Group I was 98, in Group II there were 24, in Group 1-A there were 47, and Group II-A had 66 pupils.

The mini-tests results were used to determine an estimated group mean for each of the four test groups for their performance on the complete MAT Reading Test. The estimated means were compared with the actual means for each of the groups to see if they met the criterion adopted in this study for a "good" estimate; that is, plus or minus one SEM of the population parameter.

Item data for the MAT Reading Test for individual pupils were not available. Thus it was not possible to carry out any post hoc item sampling for this test.

Summary of Methods

Use of a single item-sample. A random sample of 25 per cent was drawn proportionately from the four sections of the Stanford Arithmetic Test, Primary I, for the purpose of estimating group performance on the entire test. Both the complete test and an item-sample were administered to 153 second grade pupils in order to compare actual and estimated test parameters. The same item-sampling procedure was repeated for another group of second grade pupils using the MAT Reading Test, Primary I and II.

Context effects. Item-samples (mini-tests) were administered to pupils both as a part of a longer, more complex teacher-made arithmetic test and as mini-tests alone. Comparisons were then made of parameter estimates derived from those two empirical (a priori) sampling administration conditions and of post hoc mini-test parameter estimates obtained using the total test data base. These comparisons were made to study possible effects on pupil performance when items were administered in different contexts.

Goodness of the sampling method. Further post hoc item-sampling was carried out on the Stanford Arithmetic Test data to provide information on the effectiveness of the item-sampling approach used in this study. One hundred 25 per cent item-samples were drawn randomly from the complete test. Estimated means, standard deviations and KR21 reliability coefficients for the total test were determined for each item-sample. The frequency distributions for the estimated parameters were plotted and the standard deviation, or standard error of estimate for the sampling plan used in this study, were determined for each distribution.

Reliability. Both the estimated KR21 reliability coefficients for the complete test and the actual KR21 coefficients for the item-samples themselves were determined for both the empirical and the post hoc item-sampling data.

IV. RESULTS

Arithmetic Test

The data for the subjects taking the Stanford Achievement Test, Primary I Arithmetic Test, Form X and the mini-tests, Form X and Y were used to study the estimation of group means and standard deviations by a priori sampling as well as post hoc item-sampling. The extrapolation formulas used to calculate estimated means and standard deviations were those developed by Owens and Stufflebeam (1969). Estimated KR21 reliability coefficients were determined using a formula given by Shoemaker (1971).

The means and standard deviations for the four test groups for the complete Stanford Arithmetic Test are given in Table 3. The groups consisted of Group A which took the mini-test, Form X, Group B which took mini-test, Form Y, Group C which took mini-test, Form X within the teacher-made test, and Group D which took mini-test, Form Y within the teacher-made test. The group means for the complete test ranged from 39.48 to 53.14 and the standard deviations from 7.14 to 14.18.

Also included in Table 3 are the estimated means and standard deviations derived from the experimentally administered (a priori) mini-test results and from the post

TABLE 3

Actual and Estimated Stanford Arithmetic Test Means and Standard Deviations for Four Test Groups

Test Group	Number of Subjects	Complete Test		A Priori Mini-test		Post Hoc Mini-test	
		μ	σ	$\hat{\mu}$[a]	$\hat{\sigma}$[b]	$\hat{\mu}$	$\hat{\sigma}$
A. Mini-test, Form X	38	47.13	8.36	51.58[c]	9.16	44.34	8.07
B. Mini test, Form Y	41	39.48	14.18	50.20[c]	12.27	39.73	13.33
C. Mini-test, Form X with Teacher test	35	53.14	10.54	47.13[c]	11.25	53.22	8.79
D. Mini-test, Form Y with Teacher test	39	42.38	7.14	36.93[c]	7.17	39.38	5.53

[a]Estimated mean for complete test.

[b]Estimated standard deviation for complete test.

[c]Estimated mean not within plus or minus one SEM.

hoc mini-test results. None of the estimated means from the mini-tests actually administered to the four test groups were within plus or minus one standard error of measurement (SEM) of the actual means. The SEM for the Stanford Achievement Test, Primary I Arithmetic Test is 3.2 (Harcourt Brace Jovanovitch, Inc., 1964-1966).

Both groups of subjects taking mini-tests as a part of the teacher-made test underestimated the group means. Group C had an estimated mean 6.01 points under the actual mean, and the Group D mean was 5.45 points too low. Both groups taking only a mini-test overestimated the actual group means. Group A overestimated by 4.45 points and Group B by 10.72 points. The differences between actual and estimated standard deviations ranged from .03 to 1.91.

The four estimated means derived from the post hoc mini-test sampling data all were within one SEM of the actual means. The difference between the estimated and actual means for the four groups ranged from .08 to 3.00 points. The estimated post hoc standard deviations and the actual standard deviations differed in a range from .29 to 1.75. Thus, the post hoc item samples provided more satisfactory estimates of the group means than the estimated means derived from a priori sampling. There was little difference in the effectiveness of estimating group standard deviations by the a priori or the post hoc sampling procedures.

Actual and estimated KR21 reliability coefficients for the Stanford Arithmetic Test for each of the four groups

TABLE 4

Actual and Estimated Stanford Arithmetic Test KR21 Reliability Coefficients for Four Test Groups

Test Groups	Number of Subjects	KR21 for Complete Test	$\widehat{KR21}$[a] for A priori Mini-test	$\widehat{KR21}$ for Post hoc Mini-test
A. Mini-test, Form X	38	.87	.90	.81
B. Mini-test, Form Y	41	.95	.95	.93
C. Mini-test, Form X with Teacher test	35	.95	.92	.91
D. Mini-test, Form Y with Teacher test	39	.81	.71	.53

[a]Estimated KR21 reliability coefficient for complete test.

are given in Table 4. The range of actual KR21 coefficients for the complete Stanford Arithmetic Test was from .81 to .95. The estimated KR21 coefficients for the actual mini-tests were from .71 to .95, and for the post hoc mini-test data were from .53 to .93. Group D had the lowest actual and estimated KR21 coefficients, and also had the smallest standard deviation on both the complete test and the mini-tests.

Further Post Hoc Sampling

Post hoc sampling was carried out with the Stanford Arithmetic Test data for the total group of 153 pupils. The post hoc mini-test sample was run as well as ten additional 25 per cent random samples selected proportionately from the four sections of the Stanford Test. The total group mean and standard deviation and the estimated parameters for each of the eleven samples are presented in Appendix XV. In addition, the differences between the actual mean and the estimated means are given. The estimated means for each of the eleven post hoc samples were within one SEM of the population mean. The range of differences between the estimated means and the actual means was from .27 to 2.56 points.

The estimated standard deviations differed from the population standard deviation in a range from .07 to 1.32. Eight of the ten estimated standard deviations differed from the actual standard deviation by less than one point.

Further post hoc item-sampling was carried out on the Stanford Arithmetic Test data for the total group of pupils. Ten 25 per cent random samples (16 items) were drawn from the test as a whole without regard to the four different sections within the test. The estimated means and standard deviations for the ten samples are presented in Appendix XVI with the actual population parameters. The differences between the actual and estimated means are also given. These differences ranged from .63 to 2.40 for nine of the ten new random samples. The estimated mean for item-sample 9 was 3.54 points greater than the actual mean, a difference of more than one SEM.

The difference between the estimated standard deviations for these ten samples and the standard deviation for the complete test ranged from .40 to 2.47. Five of the ten estimated standard deviations were within one point of the actual standard deviation.

Although the differences between the estimates of means and standard deviations derived by sampling from the whole test and those obtained from stratified sampling were not great, it appeared from the data analyzed in this study that the stratified sampling approach resulted in a more efficient estimation of population parameters.

In order to provide more information for the evaluation of the effectiveness of the use of a single item-sample for estimating group parameters, still further post hoc sampling was conducted. Ninety additional 25 per cent random samples were drawn, each taken proportionately from each section of the test, making a total of 100 stratified random samples. An estimated mean, standard deviation and KR21 coefficient were determined for each sample. These estimates are presented in Appendices XVII - XIX. The estimated means ranged from 40.46 to 52.01, the standard deviation estimates from 10.05 to 13.63 and the estimated reliability coefficients from .88 to .94. The actual reliability coefficients for the 100 samples ranged from .72 to .81, with a mean of .77, and are given in Appendix XX.

The frequency distributions for the estimated means, standard deviations and KR21 coefficients for the 100 samples are shown in Figures 2 - 4. The mean of the estimated means (indicated by the broken line) was 45.39, while the actual mean for the total test was only slightly different at 45.32. The standard deviation of the estimated means, which is the standard error of estimate of the sampling method used, was 2.07. The mean of the estimated standard deviations was 11.59 and the actual standard deviation for the whole test was 11.65. The standard error of estimate for this parameter was .85. The mean of the estimated KR21 coefficients was .92 and the actual reliability for the complete test was .93. For the reliability estimates, the standard error of estimate was .03.

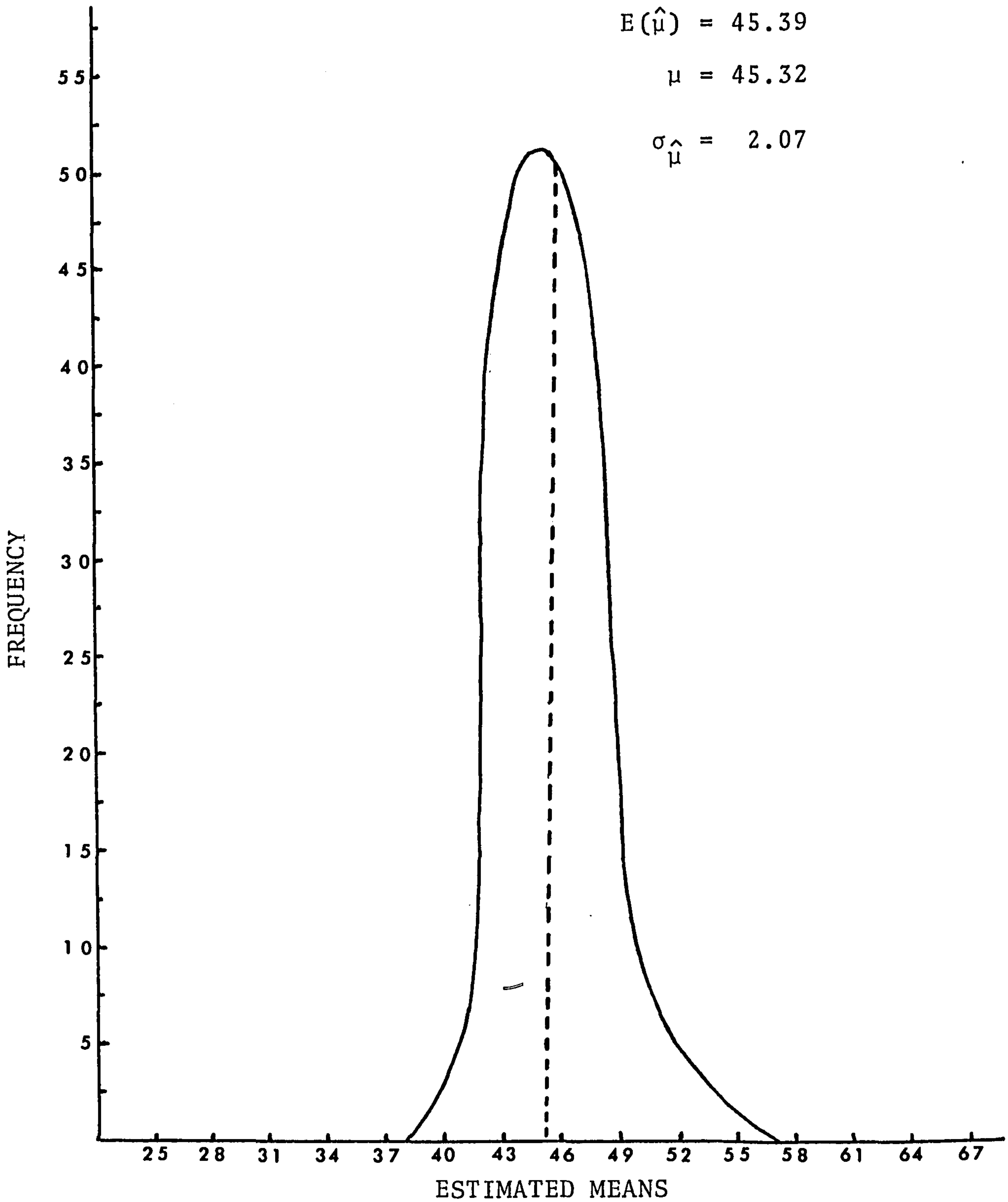


Fig. 2. Distribution of estimated population means for 100 samples.

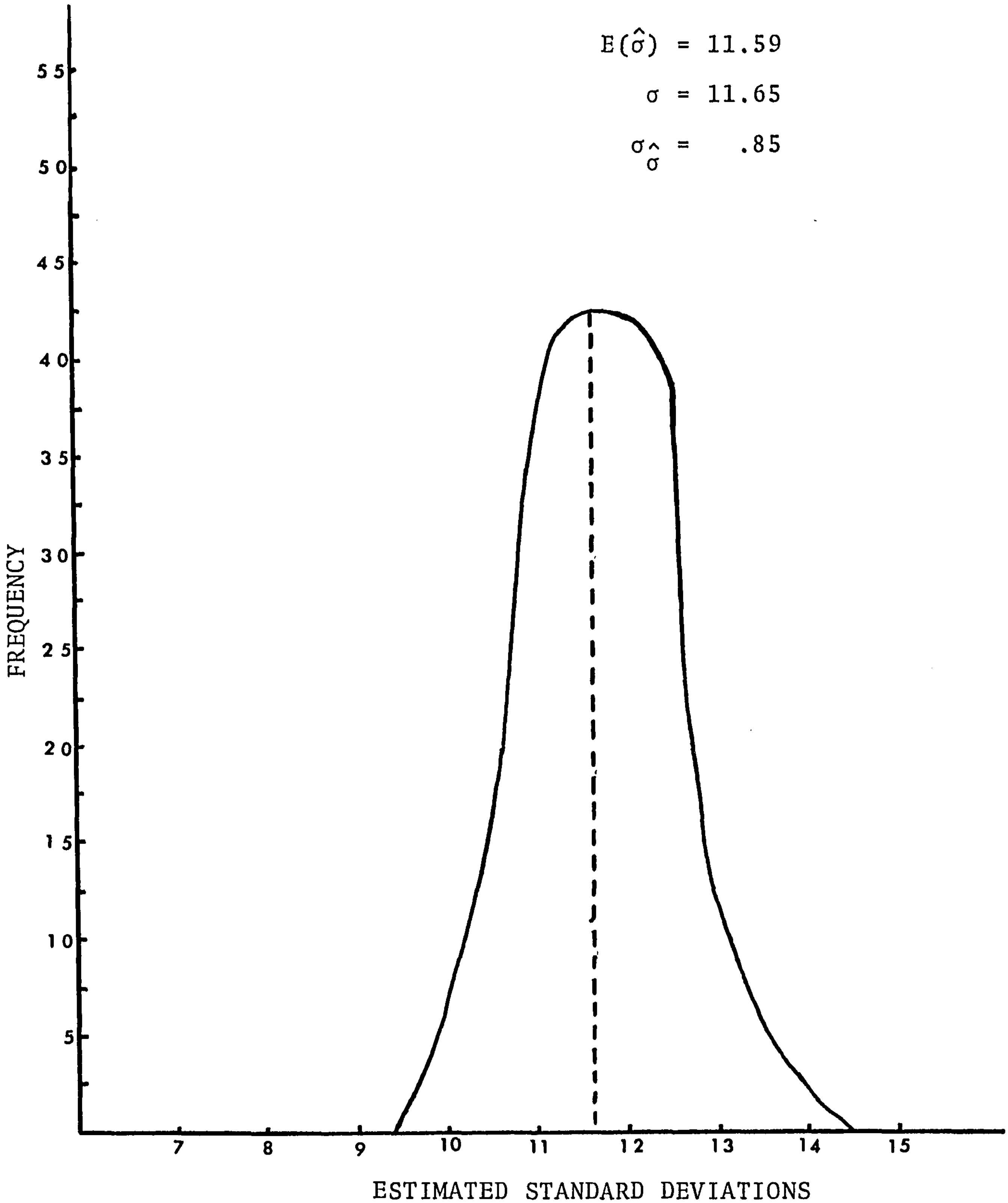


Fig. 3. Distribution of estimated population standard deviations for 100 samples.

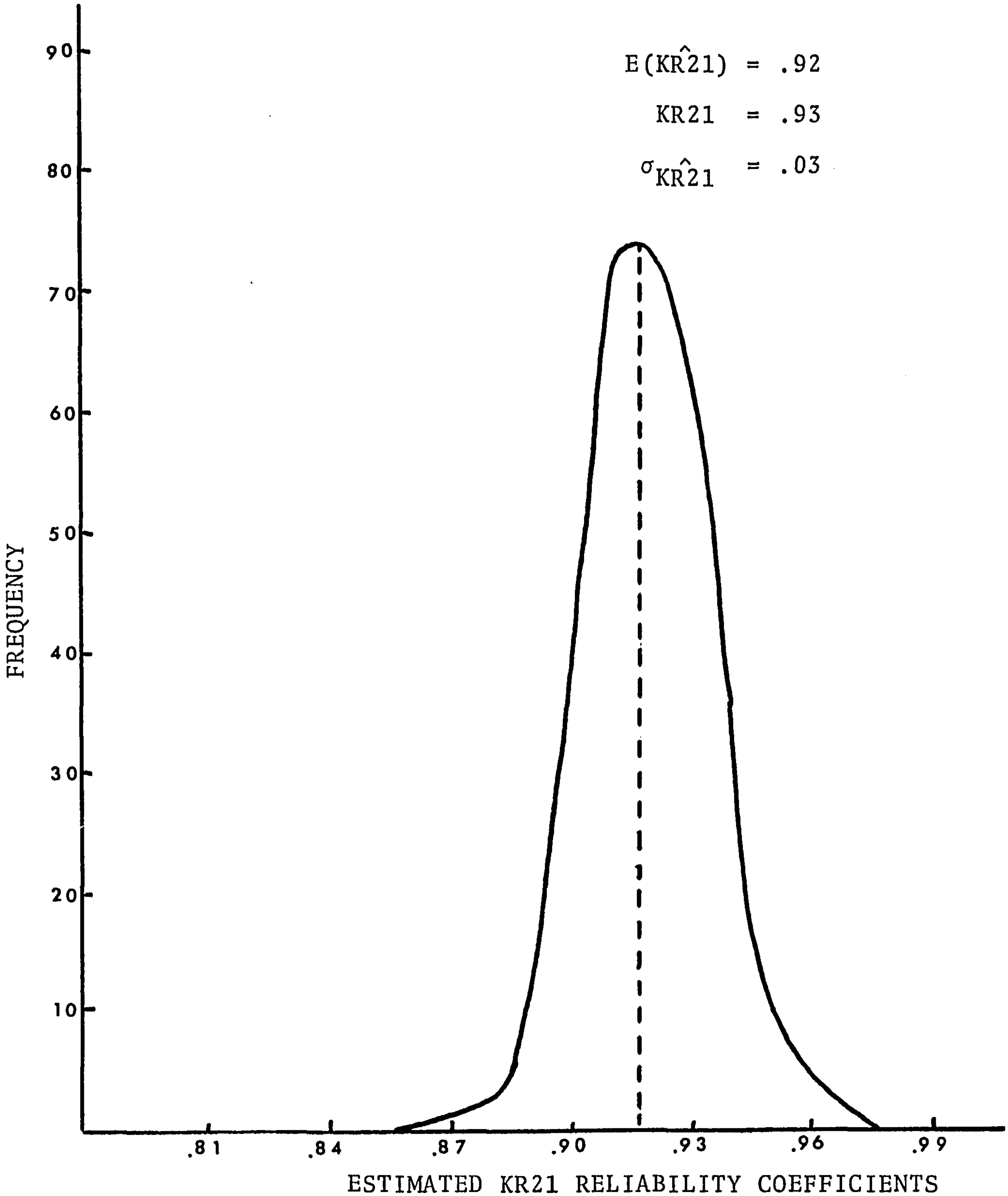


Fig. 4. Distribution of estimated KR21 reliability coefficients for 100 samples.

Reading test

The data for the four groups of pupils taking the mini-tests drawn from the MAT, Primary I and Primary II Reading Tests were used to estimate each group's mean and standard deviation for the complete test. Pupil's scores on the whole MAT Reading Test were obtained as a result of the city-wide testing program in New York City public schools, and means and standard deviations were then calculated. The four groups included Group I which took mini-test I, Group II which took mini-test II, Group III which took the mini-test I as a part of a teacher-made reading test, and Group IV which took mini-test II as a part of the same teacher-made test. Mini-test I was drawn from the Primary I reading test and mini-test II from the Primary II test.

The actual and estimated means and standard deviations for each of the four groups are given in Table 5. The estimated means for Group I, III and IV were not within one SEM of the actual means for the total test. The SEM for the MAT, Primary I Reading Test is 2.2 and for the Primary II Reading Test is 2.3 (Harcourt Brace Jovanovich, Inc., 1970). The Group II estimate of the mean did fall within one SEM of the actual mean. However, this was the smallest sample, consisting of only 24 pupils. The Group I estimate was 3.20 points higher than the actual mean, and both Group III and Group IV estimates were lower than the actual means, by 3.50 and 6.51 points respectively. These data,

TABLE 5

Actual and Estimated MAT Reading Means and Standard Deviations for Test Groups

Test Group	Number of Subjects	Complete Test		Mini-test		Complete Test	
		μ	σ	$\hat{\mu}$	$\hat{\sigma}$	KR21	$\widehat{KR21}$
I. Mini-test I	98	27.46	9.38	30.66[a]	6.81	.61	.67
II. Mini-test II	24	32.71	6.42	32.80	5.03	.36	.38
III. Mini-test I with Teacher Test	47	29.23	6.52	25.73[a]	7.79	.63	.76
IV. Mini-test II with Teacher Test	66	39.95	5.55	33.44[a]	6.19	.52	.60

[a]Estimated mean not within plus or minus one SEM.

with the exception of the small mini-test II sample, support the trend present in the Stanford Arithmetic Test data for pupils taking mini-tests only to have scores which over-estimated the group performance on the whole test, and for pupils taking mini-tests as a part of a longer, more complex teacher-made test to have scores which under-estimated the group mean on the whole test. The differences between actual and estimated standard deviations ranged from .64 to 2.57.

The actual and estimated reliability coefficients are also presented in Table 5 for each of the four test groups. Both reliability coefficients were quite low for Groups II and IV. The standard deviations for these groups were also smaller than for Groups I and III indicating greater group homogeneity, thus lower reliability.

No post hoc sampling was carried out for the reading test due to the unavailability of item data for the complete test for the subjects used in this study.

Summary of Results

Use of a single item-sample. None of the estimated means from the experimentally administered mini-tests (a priori) were within one standard error of measurement of the actual means for the Stanford Arithmetic Test. The post hoc mini-test results were all within one SEM. There was little difference between the a priori and post hoc estimates of the standard deviations. Three of the four

MAT mean estimates derived from a priori item-sampling differed from the actual means by more than one SEM. No post hoc sampling was carried out with the MAT data.

Context effects. Both groups of subjects taking Stanford Arithmetic Test mini-tests as a part of the teacher-made test underestimated the group mean on the complete test. The two groups taking only a mini-test overestimated the actual group means. The estimates from the post hoc item-samples were all closer to the actual means than the estimates from either of the a priori item-sample administration conditions. Three of the four MAT mean estimates supported the results for the Stanford mini-tests.

Goodness of the sampling method. The frequency distributions and their standard deviations (standard errors of estimate for the sampling procedure) for the means, standard deviations and KR21 coefficients estimated by the 100 post hoc random samples were determined. The means of the distributions of the estimated parameters were extremely close to the actual population values and the standard deviations were quite small, indicating that the single item-sample method was effective.

Reliability. For the 100 Stanford arithmetic post hoc mini-tests, the estimated KR21 coefficients ranged from .88 to .94 with a mean of .92. The actual reliability coefficients for the 100 mini-tests ranged from .92 to .81 with a mean of .77.

V. CONCLUSIONS AND DISCUSSION

Use of a Single Item-Sample

The item-sampling procedure used in this study, a single sample of items, was evaluated by drawing 100 post hoc samples for which the estimated means, standard deviations and reliability coefficients were determined. The standard deviations of the estimated parameters provided the standard errors of estimate for the sampling method used. The means of the sampling estimate distributions were very close to the actual parameters. The results of this sampling study indicate that a single sample of items can be used effectively in some evaluation situations. However, more research in the use of this methodology is needed, as will be discussed below.

Goodness of the Sampling Method

The use of repeated sampling to determine the standard error of estimate associated with a sampling procedure has been recommended by Shoemaker (1971b) and was an important part of this study. This approach to the evaluation of a sampling procedure would seem to be much more useful in the interpretation of the accuracy of estimates derived from sampling data than the procedures reported in other item-sampling literature. Setting a criterion for sample estimates of plus or minus one standard error of measurement for a test was utilized in the pilot study of this investigation and referred to in discussing the mini-test results. This criterion was also used by Cook and Stufflebeam (1967). However, as the standard error of measurement theoretically represents the standard deviation of an individual's observed score about his true score over many administrations of a test, it is probably more appropriate to use as a criterion for interpreting individual scores than for evaluating group scores.

The use of confidence intervals in interpreting sample estimates (Owens and Stufflebeam, 1969, Sirotnik, 1970) also involves difficulties. There is the problem of measurement error present both in the sample estimates and the actual population parameters. Both sources of error must be considered. There is the possibility that confidence bands could be determined for both the sample estimates and the population parameters and that the degree of overlap could then be used in interpreting any differences. This approach has been suggested by Feldt (1967) in another type of testing situation involving the use of confidence bands to evaluate the reliability of a difference between two scores.

Most research studies published to date in the area of item-sampling have discussed differences between sampling estimates and actual population parameters in descriptive terms without providing criteria for judging the effectiveness of the sampling estimates or the sampling procedure. It would be of value to persons considering the use of a particular sampling procedure to have information regarding the frequency distribution of parameter estimates obtained by that sampling approach and the standard error of estimate for that parameter when that sampling method is used. It would then be possible to decide whether the amount of error associated with a sampling procedure could be tolerated in the evaluation situation for which the item-sampling procedure was being selected.

Context Effects

The comparison of group performance on the complete Stanford Arithmetic Test with estimates derived from a priori and post hoc item-sampling has resulted in information useful in the consideration of possible context effects on test performance.

The overestimate of group means obtained from the mini-test results for the two groups taking only the mini-tests combined with the underestimation of total-test means for the groups taking mini-tests as a part of a teacher-made test are an indication that context effects may have influenced the subjects' performance on the mini-tests. In this study, subjects taking a mini-test of only 16 test items scored higher on those items than when they took them as a part of the original 63 item standardized test.

Pupils taking the mini-test as a part of a teacher-made test, which was longer and more complex than the original standardized test, scored lower on the items than they did when the items were administered as a part of the whole Stanford Arithmetic Test. The results of this study suggest the rather strong possibility that context effects may exist for second grade pupils who are relatively inexperienced in the art of test-taking, however the results were not conclusive. Similar results were found in a study by Cahen et al (1970) using ninth grade pupils. This indicates the possible presence of context effects for a wider age range.

The results of the testing conducted for this investigation may have been affected to some extent by the fact that the pupils in the different test groups differed in ability. This occured in spite of the fact that classes were grouped according to available information on achievement levels in an attempt to "equalize" the groups as much as possible. The data given in Table 3 which include the group means for the complete test, indicate that the grouping was only moderately successful. What effect, if any, ability or achievement level has on susceptibility to possible context effects is not known.

If a larger number of subjects had been available for this study, an elaboration of the test administration design would have been possible and would have provided further information. In addition to the four groups included in in this study, another four groups could have been selected to take the mini-tests two weeks before the complete test, rather than two weeks after. One could then determine if the order in which the testing was carried out affected the results.

Because of the possible presence of context effects indicated by the difference in results for the a priori and post hoc mini-tests, it was not possible to make any meaningful judgment regarding the use of mini-tests from the same test form as the total test as compared to the use of mini-tests from a nominally parallel form.

The results of the administration of the MAT mini-tests generally supported the trend apparent in the Stanford mini-test results regarding the possible presence of context effects on pupil performance. With the exception of one group of 24 pupils, the scores for subjects taking only the mini-test over-estimated their performance on the whole test while the scores for the two groups of pupils taking the mini-test as a part of the longer, more complex teacher-made test underestimated their total-test performance. Although the results indicate the possibility that the context in which the mini-tests were administered affected pupil performance, this must remain a possibility rather than a conclusion until further research on this subject is carried out. The results for the one test group (Group II) did not appear to be affected by context. In addition, the lack of information on performance on individual test items for the total test made it impossible to make comparisons between a priori and post hoc sampling estimates. This comparison, which was carried out for the Stanford Arithmetic Test results, would have provided more information for the assessment of the presence or absence of possible context effects.

Reliability

The use of estimated reliability coefficients which indicate what the reliability would be if the entire test were given has been questioned in this paper. It is the reliability of the item-sample itself which is most important in interpreting the results of item-samples At the least, the reliability of the item-sampling themselves should be presented, with estimated coefficients an optional consideration of doubtful value. Both sets of coefficients were presented in this study. The reliability of the 100 post hoc mini-tests ranged from .72 to .81 with a mean of .77. As mini-tests are intended solely for the purpose of group evaluation and not for the estimation of individual achievement or for decisions about individuals, the level of reliability is satisfactory. It is generally thought that the reliability of tests used for group evaluation should be at least .65 to .70 whereas the reliability necessary for evaluating individuals must be considerably higher (Mehrens and Lehmann, 1969, p. 41). Another consideration in the reliability of mini-tests or whole tests is, of course, group homogeneity. It was seen that for the Group D Stanford post hoc mini-test and the Group II and Group IV MAT mini-tests the reliability coefficients were lower than is acceptable. These groups also had low standard deviations. One possible solution to this problem is to use larger groups of subjects thus reducing the likelihood of having groups which are too homogeneous. When the total group of 153 subjects were used for the analysis of the 100 post hoc samples the problem of group homogeneity was solved.

Suggestions for Future Study

The results of this study, as well as problems encountered in carrying out the study, have shown that there are a number of areas worthy of future investigation. These suggested areas of study are related both to the use of a single item-sample and to matrix-sampling.

One important research area is the effect of population size on the accuracy of item-sample estimates of group parameters. It is possible that the use of item-sampling might be suitable for estimating performance in a school system, a school or for a grade level composed of several classes. However, the same sampling approach may not be as effective in the evaluation of the performance of a single class or a small group of pupils due to increased errors of estimation. The problem of group homogeneity discussed above is related to the issue of population size.

Although selection of items by the use of random sampling is the simplest procedure and generally the most practical, particularly when item analysis information for a test is lacking or difficult to obtain, it may be that other criteria for item selection would result in more efficient item-sampling estimates of group performance. In a study of multiple matrix-sampling versus examinee-sampling by Barcikowski (1972), items were selected according to their range of difficulty and range of biserial correlation. Such criteria may also prove effective in selecting items for a single item-sample approach.

Perhaps the most important area for further investigation is that of possible context effects on pupil performance when samples of items are administered rather than complete tests. Although the results of this study are not conclusive on this issue, they do support the possibility that performance on test items is affected by the context in which the items are presented, particularly for young pupils. Further information on the relationship, if any, between the context in which items are presented and the effect upon examinees of different age levels is needed. It would also be helpful to know more about the presence or absence of context effects for item-samples drawn from a wider variety of tests than the reading and arithmetic tests utilized in this study. Finally, the investigation of the relationship between the size of the item-sample administered and the possible context effects on performance would also be useful. The results of this study indicate that young pupils may perform better when given a small number of items than when given a longer test. If this is true, then it could be that pupil performance would be improved, for some populations, if a test is administered in the form of numerous mini-tests over a period of time rather than as a complete test. This could be investigated by administering a complete test, which has been designed to be given in one test period, to a group of pupils, and, by then administering the test to the same group as a series of item-samples over a longer period of time. A comparison of the test scores obtained by the two administration procedures would provide further evidence regarding the effects of context on test performance.

One would expect that there is a direct linear relationship between the size of item-sample selected and the accuracy of estimation of group performance on the complete test. However, it could be worthwhile to test this assumption through the selection of a wider range of sizes of item-samples than the four sample sizes utilized in the pilot study for the investigation reported here. Information on the sampling error associated with different sizes of item-samples would be helpful to persons considering the use of this sampling procedure.

The size of item-sample required to estimate group performance is related to the degree of homogeneity of items found in standardized tests. An investigation of this area may reveal that items included in standardized tests are less heterogeneous than popularly assumed. The fact that a single sample of items from such a test could provide a fairly good estimate of group performance on the total test, as in this study, indicates that this is true or at least a real possibility.

Summary

The results of this investigation suggested that the use of a single sample of items for the estimation of group performance may prove to be an effective approach to item-sampling and more practical than the traditional multiple matrix-sampling approach. The general assumption that performance on item-samples is not affected by their context has been placed in question. Further research is recommended to provide more information about the use of a single sample of items for estimating group parameters and to shed more light upon the issue of context effect.

Appendix I

Arithmetic Mini-Test, Form X

II. *Part A: Measures* (X)

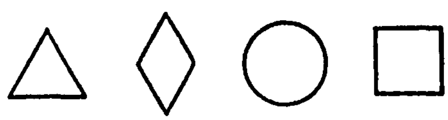
SAMPLES

A

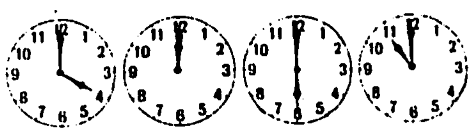
1 ~~2~~ 3 4

B

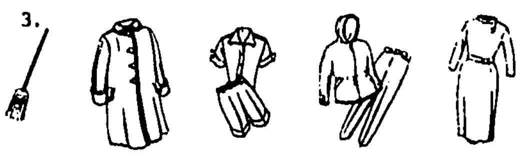
1.



2.



3.



GO ON TO THE NEXT PAGE ▶

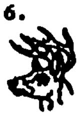
II. ***Part B: Problem Solving*** (x)

4.  2 3 5 6

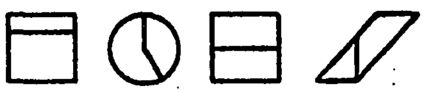
7.  1 2 3 4

5.  1 2 3 10

8.  1 3 7 10

6.  1 4 5 10

Part C: Number Concepts

9.  

10.  ______

11.  159 □

12. $\begin{array}{r} 6 \\ +3 \\ \hline \end{array}$

13. $\begin{array}{r} 5 \\ 3 \\ +3 \\ \hline \end{array}$

14. $\begin{array}{r} 9 \\ -8 \\ \hline \end{array}$

15. $\begin{array}{r} 7 \\ -4 \\ \hline \end{array}$

16. □ − 2 = 2

STOP

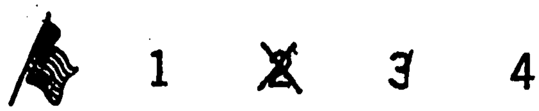
Appendix II

Arithmetic Mini-Test, Form Y

II. ***Part A: Measures*** (Y)

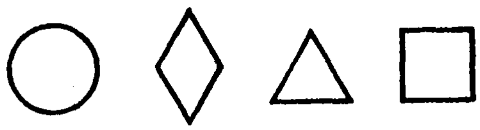
SAMPLES

A

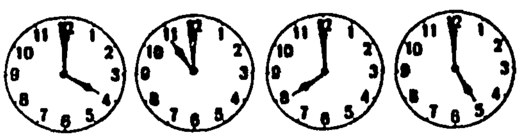


B

1.



2.



3.

GO ON TO THE NEXT PAGE ▶

II. *Part B: Problem Solving* (Y)

4.  4 10 14 40

5.  1 2 3 5

6.  5 10 20 40

7.  2 3 4 6

8.  1 2 4 8

Part C: Number Concepts

9.  2 3 4 5

10.  ______

11.  256 ☐

12. $\begin{array}{r} 7 \\ +3 \\ \hline \end{array}$

13. $\begin{array}{r} 2 \\ 4 \\ +5 \\ \hline \end{array}$

14. $\begin{array}{r} 8 \\ -7 \\ \hline \end{array}$

15. $\begin{array}{r} 8 \\ -3 \\ \hline \end{array}$

16.  ☐ − 4 = 1

STOP

Appendix III

Teacher-Made Arithmetic Test

I. Put an "X" by the correct answer.

SAMPLE:
```
  2
+ 4
---
```

_____ 2

_____ 5

__x__ 6

_____ 8

---

1. 7 + 2

_____ 5

_____ 8

_____ 9

_____ 14

2. 54 + 3

_____ 41

_____ 51

_____ 57

_____ 67

3. 6 + ___ = 6

_____ 0

_____ 1

_____ 6

_____ 12

4. 84 + 25 = ___

_____ 69

_____ 108

_____ 109

_____ 110

5. 54 + 21 + 32

_____ 98

_____ 99

_____ 107

_____ 108

6. $5.23 + 4.56

_____ $1.33

_____ $1.77

_____ $9.78

_____ $9.79

7. 15 - ___ = 8

_____ 7

_____ 8

_____ 23

_____ 24

8. 29 - ___ = 20

_____ 8

_____ 9

_____ 20

_____ 49

9. ___ - 3 = 11

_____ 8

_____ 14

_____ 15

_____ 33

10. $10.82
- 2.51

_____ $ 8.30

_____ $ 8.31

_____ $12.32

_____ $12.43

11. 4 X ___ = 400

_____ 0

_____ 4

_____ 40

_____ 100

12. 10 X 10 = ___

_____ 0

_____ 10

_____ 20

_____ 100

13. 2 X 8 = ___

_____ 6

_____ 10

_____ 14

_____ 16

14. 5 X 9 = ___

_____ 4

_____ 14

_____ 40

_____ 45

15. $\frac{1}{2}$ X 18 = ___

_____ 6

_____ 9

_____ 20

_____ 36

STOP!

III. Put an "X" by the correct answer.

SAMPLE: 1, 2, 3, 4, 5, 6, ___.

The next number is:

_____ 5

_____ 6

__x__ 7

_____ 8

---

1. 2, 4, 6, 8, 10, 12, 14, ___.

The next number is:

_____ 15

_____ 16

_____ 18

_____ 24

2. 5, 10, 15, 20, 25, 30, ___, ___.

_____ 31, 32

_____ 35, 40

_____ 40, 45

_____ 40, 50

3. 10, 20, 30, 40, 50, ___, ___, ___.

The next three numbers are:

_____ 35, 45, 55

_____ 50, 60, 70

_____ 55, 60, 65

_____ 60, 70, 80

4. Which is the missing number?

25, ___, 35, 40

_____ 26

_____ 28

_____ 30

_____ 32

5. Which is the missing number?

90 ___ 70 60

_____ 75

_____ 80

_____ 85

_____ 100

6. Which number is missing?

(4 + 2) + 11 = 4 + (___ + 11)

_____ 2

_____ 5

_____ 6

_____ 11

IV. Listen carefully to the question and put an "X" by the correct answer.

SAMPLE: ______ 2
______ 20
______ 180
______ 200

---

1. ______ 1
______ 13
______ 14
______ 15

2. ______ 7
______ 8
______ 9
______ 13

3. ______ 2
______ 15
______ 20
______ 80

4. ______ 0
______ 1
______ 2
______ 3

5. ______ 4,000
______ 4,005
______ 4,050
______ 4,500

6. ______ 3,000
______ 3,060
______ 3,700
______ 3,706

7. ______ 1
______ 11
______ 25
______ 30

8. ______ 2
______ 8
______ 15
______ 20

9. ______ 4
______ 6
______ 10
______ 16

10. ______ 0
______ 8
______ 10
______ 80

STOP!

V. Read each problem and put an "X" by the correct answer.

1. Sam has 7 books. He gives Tom 2 books. How many books does he have left?

_____ 0

_____ 3

_____ 5

_____ 9

2. My sister ate 3 pieces of candy. She wants to eat 10 pieces of candy altogether. How many more pieces of candy will she eat?

_____ 3

_____ 7

_____ 10

_____ 13

3. A bag of potato chips costs ten cents. Jack has fifty cents. How many bags of potato chips can he buy?

_____ 5

_____ 10

_____ 15

_____ 50

4. Our class can sing 10 songs in one hour. How many songs can we sing in five hours?

_____ 20

_____ 50

_____ 100

_____ 500

5. Pat bought a coloring book for 20¢, crayons for 25¢, and some gum for 10¢. How much money did she spend altogether?

_____ 15¢

_____ 50¢

_____ 55¢

_____ 60¢

6. There are 30 children in our class when everyone is here. If 10 children are absent, how many are here?

_____ 10

_____ 20

_____ 30

_____ 40

7. Mom gives Pat 5¢ every time she gets a new tooth. How many new teeth does Pat have to get for her to get 25¢?

_____ 1

_____ 2

_____ 5

_____ 25

8. If I divide 8 pieces of pie equally between two boys, how many pieces will each boy get?

_____ 1

_____ 2

_____ 4

_____ 8

STOP!

Appendix IV

Administration Instructions - Second Grade

Arithmetic Mini-Test, Form X*

Test II (Time: about 9 minutes)

Say to the pupils:

Turn your booklets to page 3. (See that they all do this correctly.) Find the picture of the flag near the top of the page. Put your finger on the flag. Look at the numbers beside the picture of the flag. There is a cross on one of the numbers. Which number is it? (Pause for the class to answer.) Yet, it is the number 2.

Now look at the pictures of the flowers in the next row. Make a cross on the tallest flower. (Pause.) Have you made a cross on the tallest flower?

Check to make sure that each pupil has marked this sample correctly.

Read each item number, then read each item exactly as it is stated and repeat it once. Pause about 10 seconds before going on to the next item. The total time necessary to administer this test is approximately 9 minutes.

Part A: Measures

1. Now look at the four figures or shapes. Draw a line from the square to the circle.

2. See the next row. Look at the clocks. Make a cross on the clock that shows what time it is when it is noon.

3. See the broom. Look at the pictures of the articles of women's clothing. Make a cross on what a girl would most likely wear outside when the temperature is about 85°.

Ask the pupils to turn their booklets to page 4. Be sure they have done this.

Part B: Problem Solving

4. Now see the horse's head and the numbers beside it. Daddy has 3 letters to mail. He must put 2 stamps on each of them. How many stamps does he need in all? Make a cross on the number.

5. Look at the numbers beside the dog. Daddy's car will hold 6 people. Aunt Mary's car will hold 4 people. Daddy's car will hold how many more people than Aunt Mary's car will hold? Make a cross on the number.

6. Find the deer's head and the numbers after it. Bill wants a toy that costs 50¢. He saves 10¢ each week. How many weeks will it take him to save enough money to buy the toy? Make a cross on the number of weeks.

7. See the numbers beside the chicken at the top of the page. A box of candy and a bag of marbles weight 5 pounds together. The box of candy weights 3 pounds. How many pounds does the bag of marbles weigh? Make a cross on the number.

8. Find the bicycle and the numbers beside it. The children hiked 7 miles in 7 hours. At this rate, how many miles did they hike in 10 hour? Make a cross on the number of miles.

Part C: Number Concepts

9. Look at the figures beside the balloon. Make a cross on the figure that is divided into halves.

10. Find the glass and the line after it. Write the number 74 on this line.

11. Look at the fork. See the 1, the 5, and the 9. Which of them is in the hundreds' place? Write the number in the little box.

Pause.

The rest of the examples on this page are addition and subtraction examples. You must watch the signs carefully and do what they tell you. Do not write anything until I tell you to do so.

Look at the first row where you see the button. Put your finger on the button. You will add in the first two examples in this row. In the last two examples in this row you will subtract; you will take away.

Now put your finger on the show. Look at the board and I will show you how to do the last problem on the page.

NOTE: Write on the chalkboard the example: $2 + \square = 3$

Say to the children:

Two plus what number equals three? Yes, two plus one equals three, so I write a one in the box. Be sure you read the sign for the example so that you will know what to write in the box.

Now go back to where the button is. Do the first row and then go right on to the last example next to the shoe.

READY, GO!

At the end of two minutes, say: STOP! Close your booklet and put your pencil down.

Appendix V

Administration Instructions - Second Grade

Arithmetic Mini-Test, Form Y*

Test II (Time: about 9 minutes)

Say to the pupils:

Turn your booklets to page 3. (See that they all do this correctly.) Find the picture of the flag near the top of the page. Put your finger on the flag. Look at the numbers beside the picture of the flag. There is a cross on one of the numbers. Which number is it? (Pause for the class to answer.) Yes, it is the number 2.

Now look at the pictures of the flowers in the next row. Make a cross on the tallest flower. (Pause.) Have you made a cross on the tallest flower?

Check to make sure that each pupil has marked this sample correctly.

Read each item number, then read each item exactly as it is stated and repeat it once. Pause about 10 seconds before going on to the next item. The total time necessary to administer this test is approximately 9 minutes.

Part A: Measures

1. Now look at the four figures or shapes in the next row. Draw a line from the square to the triangle.

2. Look at the next row. See the clocks. Make a cross on the clock that shows when many people eat breakfast.

3. See the broom. Look at the pictures of the things next to it. Each one has a thermometer in it. Make a cross on the picture in which the temperature would be about 200°.

Ask the pupils to turn their booklets to page 4. Be sure they have done this.

Part B: Problem Solving

4. Now see the horse's head and the numbers beside it. If 1 candy bar costs 10¢, how many cents will it take to buy 4 candy bars? Make a cross on the number.

5. Look at the numbers beside the dog. Daddy planted 2 bushes. One of them is 3 feet tall. The other is 2 feet tall. One bush is how many feet taller than the other? Make a cross on the number.

6. Find the deer's head and the numbers after it. You have 30¢ to spend. You buy 3 candy apples with it. Each candy apple costs the same amount. How many cents does each candy apple cost? Make a cross on the number of cents.

7. See the numbers beside the chicken at the top of the next column. There are 6 children in Mary's reading group when everyone is present. Today 3 children are present. How many children are not present? Make a cross on the number.

8. Find the bicycle and the numbers beside it. Philip hiked 4 miles in 4 hours. At this rate, how many miles did he hike in 1 hour? Make a cross on the number of miles.

Part C: Number Concepts

9. Look at the numbers beside the balloon. When we divide something into thirds, how many pieces are there? Make a cross on the number of pieces there are.

10. Find the glass and the line after it. Write the number 79 on this line.

11. Look at the fork. See the 2, the 5, and the 6. Which of them is in the tens' place? Write the number in the little box.

Pause.

The rest of the examples on this page are addition and subtraction examples. You must watch the signs carefully and do what they tell you. Do not write anything until I tell you to do so.

Look at the first row where you see the button. Put your finger on the button. You will add in the first two examples in this row. In the last two examples in this row you will subtract; you will take away.

Now put your finger on the shoe. Look at the board and I will show you how to do the last problem on the page.

NOTE: Write on the chalkboard the example: 2 + ☐ = 3

Say to the children:

Two plus what number equals three? Yes, two plus one equals three, so I write a one in the box. Be sure you read the sign for the example so that you will know what to write in the box.

Now go back to where the button is. Do the first row and then go right on to the last example next to the shoe.

READY, GO!

At the end of two minutes, say: STOP! Close your booklet and put your pencil down.

Appendix VI

Administration Instructions - Teacher-Made Arithmetic Test with Mini-Test, Form X*

Be sure that each pupil's name, class, and the date are written on the cover of his test booklet.

Test I (Time: about 10 minutes)

Ask the pupils to turn to page 1 of their booklets.

Read the sample problem to them: 2 + 4 is ___, and point out that an "X" has been placed next to the correct answer: 6.

Ask if there are any questions. Then ask the pupils to begin with problem 1 and to work all of the problems on page 1 and page 2, marking an "X" by the correct answer for each problem.

When they finish page 2, they should stop and put their pencils down.

* * * * * * * * * * *

Test II (Time: about 9 minutes)

Say to the pupils:

Turn your booklets to page 3. (See that they all do this correctly.) Find the picture of the flag near the top of the page. Put your finger on the flag. Look at the numbers beside the picture of the flag. There is a cross on one of the numbers. Which number is it? (Pause for the class to answer.) Yes, it is the number 2.

Now look at the pictures of the flowers in the next row. Make a cross on the tallest flower. (Pause.) Have you made a cross on the tallest flower?

Check to make sure that each pupil has marked this sample correctly.

Read each item number, then read each item exactly as it is stated and repeat it once. Pause about 10 seconds before going on to the next item. The total time necessary to administer this test is approximately 9 minutes.

*Test II administration instructions reproduced from Stanford Achievement Test, Teacher's Directions, copyright

Part A: Measures

1. Now look at the four figures or shapes. Draw a line from the square to the circle.

2. See the next row. Look at the clocks. Make a cross on the clock that shows what time it is when it is noon.

3. See the broom. Look at the pictures of the articles of women's clothing. Make a cross on what a girl would most likely wear outside when the temperature is about 85°.

Test II (cont.)

Ask the pupils to turn their booklets to page 4. Be sure they have done this.

Part B: Problem Solving

4. Now see the horse's head and the numbers beside it. Daddy has 3 letters to mail. He must put 2 stamps on each of them. How many stamps does he need in all? Make a cross on the number.

5. Look at the numbers beside the dog. Daddy's car will hold 6 people. Aunt Mary's car will hold 4 people. Daddy's car will hold how many more people than Aunt Mary's car will hold? Make a cross on the number.

6. Find the deer's head and the numbers after it. Bill wants a toy that costs 50¢. He saves 10¢ each week. How many weeks will it take him to save enough money to buy the toy? Make a cross on the number of weeks.

7. See the numbers beside the chicken at the top of the page. A box of candy and a bag of marbles weigh 5 pounds together. The box of candy weighs 3 pounds. How many pounds does the bag of marbles weigh? Make a cross on the number.

8. Find the bicycle and the numbers beside it. The children hiked 7 miles in 7 hours. At this rate, how many miles did they hike in 1 hour? Make a cross on the number of miles.

Part C: Number Concepts

9. Look at the figures beside the balloon. Make a cross on the figure that is divided into halves.

10. Find the glass and the line after it. Write the number 74 on this line.

11. Look at the fork. See the 1, the 5, and the 9. Which of them is in the hundreds' place? Write the number in the little box.

Pause.

The rest of the examples on this page are addition and subtraction examples. You must watch the signs carefully and do what they tell you. Do not write anything until I tell you to do so.

Look at the first row where you see the button. Put your finger on the button. You will add in the first two examples in this row. In the last two examples in this row you will subtract; you will take away.

Now put your finger on the shoe. Look at the board and I will show you how to do the last problem on the page.

Test II: Part C (cont.)

NOTE: Write on the chalkboard the example $2 + \square = 3$

Say to the children:

Two plus what number equals three? Yes, two plus one equals three, so I write a one in the box. Be sure you read the sign for the example so that you will know what to write in the box.

Now go back to where the button is. Do the first row and then go right on to the last example next to the shoe.

READY, GO!

At the end of two minutes, say: STOP! Close your booklet and put your pencil down.

* * * * * * * * * *

Test III (Time: about 4 minutes)

Ask the pupils to turn to page 5.

Say: In the sample problem at the top of the page, there is a series of numbers: 1, 2, 3, 4, 5, 6, followed by a line. The problem asks what the next number is. There is an "X" next to 7 which is the correct answer.

In each example on this page, see how the numbers go and put an "X" by the answer that tells you the number of numbers that go on the line.

Are there any questions? (Give further instructions if needed.)

When you finish this page, put your pencils down.

* * * * * * * * * * *

At this point the pupils may take a short rest period or you may decide to complete Tests IV and V at another time.

Test IV (Time: about 5 minutes)

Ask the pupils to turn to page 6.

Say: For each of the examples on this page, I will read the question to you and you are to put an "X" next to the correct answer. Look at the sample problem at the top of the page. Listen while I read the question to you....100 minus 80 equals what? (Ask the class for the correct answer and see that they put an "X" by that answer.) Now listen carefully as I read the rest of the questions for this page. Put an "X" by the correct answer for each question that I read.

1. 6 + 7 = ___
2. 2 + 5 + 6 = ___
3. 50 - 30 = ___
4. 82 - 81 = ___
5. 4005 - 5 = ___
6. 3760 - 760 = ___
7. 6 x 5 = ___
8. 5 x 3 = ___
9. 1/2 x 8 = ___
10. How many 10's are there in 80?

* * * * * * * * * * *

Test V (Time: about 10 minutes)

Say: Turn to page 7. Read each problem on this page and put and "X" by the correct answer.

(Some of these problems may be difficult for some second grade children. Encourage them to do their best, but to skip those problems which they don't understand.)

* * * * * * * * * * *

Thank you!

Appendix VII

Administration Instructions - Teacher-Made Arithmetic Test with Mini-Test, Form Y*

Be sure that each pupil's name, class, and the date are written on the cover of his test booklet.

Test I (Time: about 10 minutes)

Ask the pupils to turn to page 1 of their booklets.

Read the sample problem to them: 2 + 4 is ___, and point out that an "X" has been placed next to the correct answer, 6.

Ask if there are any questions. Then ask the pupils to begin with problem 1 and to work all of the problems on page 1 and page 2, marking an "X" by the correct answer for each problem.

When they finish page 2, they should stop and put their pencils down.

* * * * * * * * * * *

Test II (Time: about 9 minutes)

Say to the pupils:

Turn your booklets to page 3. (See that they all do this correctly.) Find the picture of the flag near the top of the page. Put your finger on the flag. Look at the numbers beside the picture of the flag. There is a cross on one of the numbers. Which number is it? (Pause for the class to answer.) Yes, it is the number 2.

Now look at the pictures of the flowers in the next row. Make a cross on the tallest flower. (Pause.) Have you make a cross on the tallest flower?

Check to make sure that each pupil has marked this sample correctly.

Read each item number, then read each item exactly as it is stated and repeat it once. Pause about 10 seconds before going on to the next item. The total time necessary to administer this test is approximately 9 minutes.

Part A: Measures

1. Now look at the four figures or shapes in the next row. Draw a line from the square to the triangle.

2. Look at the next row. See the clocks. Make a cross on the clock that shows when many people eat breakfast.

3. See the broom. Look at the pictures of the things next to it. Each one has a thermometer in it. Make a cross on the picture in which the temperature would be about 200°.

Ask the pupils to turn their booklets to page 4. Be sure they have done this.

Part B: Problem Solving

4. Now see the horse's head and the numbers beside it. If 1 candy bar costs 10¢, how many cents will it take to buy 4 candy bars? Make a cross on the number.

5. Look at the numbers beside the dog. Daddy planted 2 bushes. One of them is 3 feet tall. The other is 2 feet tall. One bush is how many feet taller than the other? Make a cross on the number.

6. Find the deer's head and the numbers after it. You have 30¢ to spend. You buy 3 candy apples with it. Each candy apple costs the same amount. How many cents does each candy apple cost? Make a cross on the number of cents.

7. See the numbers beside the chicken at the top of the next column. There are 6 children in Mary's reading group when everyone is present. Today 3 children are present. How many children are not present? Make a cross on the number.

8. Find the bicycle and the numbers beside it. Philip hiked 4 miles in 4 hours. At this rate, how many miles did he hike in 1 hour? Make a cross on the number of miles.

Part C: Number Concepts

9. Look at the numbers beside the balloon. When we divide something into thirds, how many pieces are there? Make a cross on the number of pieces there are.

10. Find the glass and the line after it. Write the number 79 on this line.

11. Look at the fork. See the 2, the 5, and the 6. Which of them is in the tens' place? Write the number in the little box.

Pause.

The rest of the examples on this page are addition and subtraction examples. You must watch the signs carefully and do what they tell you. Do not write anything until I tell you to do so.

Look at the first row where you see the button. Put your finger on the button. You will add in the first two examples in this row. In the last two examples in this row you will subtract; you will take away.

Now put your finger on the shoe. Look at the board and I will show you how to do the last problem on the page.

NOTE: Write on the chalkboard the example: 2 + □ = 3

Say to the children:

Two plus what number equals three? Yes, two plus one equals three, so I write a one in the box. Be sure you read the sign for the example so that you will know what to write in the box.

Now go back to where the button is. Do the first row and then go right on to the last example next to the shoe.

READY, GO!

At the end of two minutes, say: STOP! Close your booklet and put your pencil down.

* * * * * * * * * * *

Test III (Time: about 4 minutes)

Ask the pupils to turn to page 5.

Say: In the sample problem at the top of the page, there is a series of numbers: 1, 2, 3, 4, 5, 6, followed by a line. The problem asks what the next number is. There is an "X" next to 7 which is the correct answer.

In each example on this page, see how the numbers go and put an "X" by the answer that tells you the number or numbers that go on the line.

Are there any questions? (Give further instructions if needed.)

When you finish this page, put your pencils down.

* * * * * * * * * * *

At this point the pupils may take a short rest period or you may decide to complete Tests IV and V at another time.

Test IV (Time: about 5 minutes)

Ask the pupils to turn to page 6.

Say: For each of the examples on this page, I will read the question to you and you are to put an "X" next to the correct answer. Look at the sample problem at the top of the page. Listen while I read the question to you.... 100 minus 80 equals what? (Ask the class for the correct answer and see that they put an "X" by that answer.) Now listen carefully as I read the rest of the questions for this page. Put an "X" by the correct answer for each question that I read.

1. 6 + 7 = ___
2. 2 + 5 + 6 = ___
3. 50 - 30 = ___
4. 82 - 81 = ___
5. 4005 - 5 = ___
6. 3760 - 760 = ___
7. 6 x 5 = ___
8. 5 x 3 = ___
9. 1/2 x 8 = ___
10. How many 10's are there in 80?

* * * * * * * * * * *

Test V (Time: about 10 minutes)

Say: Turn to page 7. Read each problem on this page and put an "X" by the correct answer.

(Some of these problems may be difficult for some second grade children. Encourage them to do their best, but to skip those problems which they don't understand.)

* * * * * * * * * * *

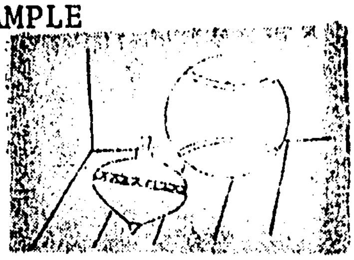
Thank you!

Appendix VIII

Reading Mini-Test I

III: A: Sentences

SAMPLE

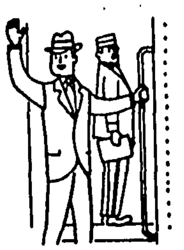


- The toys are on the table.
- These toys are for children.
- The top is bigger than the ball.

1.

- Mother and Ann are riding their bicycles.
- Mother and Ann are driving to Aunt Ellen's.
- Mother is taking Ann to school.

2.



- The man is getting on the train.
- The train is moving fast.
- The train has a flat tire.

3.

- The boy watches the sun.
- The boy looks out at the starry night.
- Ted pulls down the window shade.

STOP

III: B: Stories

SAMPLE: I am like a short coat.

I keep you warm.

I am a ____

⬭ jacket ⬭ raincoat ⬭ dress

4.

I am very sad.

Tears come from my eyes.

What am I doing?

⬭ crying ⬭ laughing ⬭ frowning

5.

I am part of your body.

I do not make a noise.

I can hear sounds.

What am I?

⬭ eyes ⬭ mouth ⬭ ears

STOP

III: B Stories (Continued)

SAMPLE:

Ted has a puppy. His name is Happy. Ted and Happy like to play.	A The pet is a ____ ◯ dog ◯ boy ◯ toy B Ted and Happy like to ____ ◯ play ◯ work ◯ look
We go to school by bus. It is a special yellow school bus. Today we almost missed the bus. We had to run to catch it.	6. We ride to school on a ___ ◯ bus ◯ bicycle ◯ scooter 7. Today we were ____ ◯ sick ◯ almost late ◯ on vacation 8. We caught the bus by ____ ◯ hopping ◯ laughing ◯ running
Mary went to the store. She wanted to buy chocolates. She saw green candy. Then she saw red candy. Then she saw her brown candy. She gave the store man her money. She took her favorite candy home.	9. What color did she see first? ◯ white ◯ green ◯ red 10. Chocolate candy is ____ ◯ green ◯ red ◯ brown 11. Mary gave the store man her ____ ◯ money ◯ hand ◯ candy

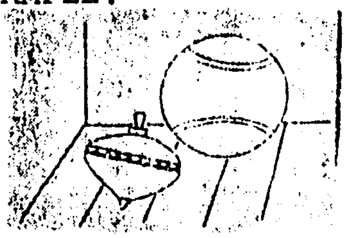
STOP

Appendix IX

Reading Mini-Test II

III: A: Sentences

SAMPLE:

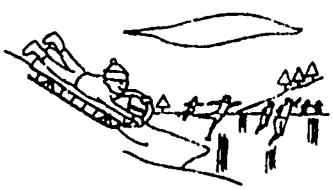


- These toys are for children.
- The toys are on the table.
- The top is bigger than the ball.

1.

- The fireman is doing some painting.
- The fireman has the tools for putting out a fire.
- A fireman works by himself.

2.



- The ice is not strong enough to hold the skaters.
- The pond is still frozen, but all the snow has melted.
- One boy brought his sled instead of his skates.

3.

- It is colder inside the house than outdoors.
- Most people would live in this house without heat.
- This house must be heated in winter.

STOP

III: B: Stories

SAMPLE:

Mother made a cake.
She put candles on it.
The candles told how old Jane was.

A. Mother was getting ready for ___

- ⬭ a birthday
- ⬭ summer
- ⬭ Thanksgiving

B On the cake, Mother put ___

- ⬭ candles
- ⬭ candy
- ⬭ ice cream

Once upon a time some soldiers played a trick to get into a city with walls. They built a wooden horse. It was as tall as a mountain. A hundred soldiers climed inside it. Some men rolled it to the wall and left it. The people in the city thought it was a gift. When they took the horse inside the walls, out came the soldiers.

4. Around the city was a ____
   ⬭ forest ⬭ river ⬭ wall

5. The horse was made of ____
   ⬭ gold ⬭ silver ⬭ wood

6. Why did the soldiers build it?
   - ⬭ to get into the city
   - ⬭ to make people laugh
   - ⬭ to ride it

7. What did they pretend the horse was?
   ⬭ a gift ⬭ a dog ⬭ a fence

2

Mike's tenth birthday present was a shiny red bicycle. At first, he couldn't ride it. No matter how many times he tried, he always fell off. In the afternoon, Mike's brother, Bill, said he would help him learn to ride. Bill gave Mike lessons on a quiet side street. Mike fell off the bike several more times, but he learned to ride that day.

8. The best name for this story is ____
   - ⬭ Mike's Birthday
   - ⬭ Mike's Brother
   - ⬭ Mike's New Bicycle

9. Mike's brother is probably ____
   - ⬭ handsome
   - ⬭ kind
   - ⬭ old

10. Bill gave Mike lessons on a side street because ____
   - ⬭ there was no traffic
   - ⬭ there were many children there
   - ⬭ the bike was new

11. Mike learned to ride a bike in ____
   - ⬭ a day
   - ⬭ a week
   - ⬭ one morning

STOP

Appendix X

Teacher-Made Reading Test

I. Listen carefully and mark an "X" by the word I say.

Sample:

__x_ sing

____ sting

____ song

____ sin

1. ____ see
   ____ sleep
   ____ slept
   ____ sheep

2. ____ shell
   ____ chin
   ____ ship
   ____ cherry

3. ____ sand
   ____ send
   ____ stand
   ____ spend

4. ____ high
   ____ thigh
   ____ this
   ____ night

5. ____ house
   ____ horse
   ____ home
   ____ hate

6. ____ chill
   ____ child
   ____ children
   ____ chime

7. ____ money
   ____ mud
   ____ must
   ____ much

8. ____ elephant
   ____ examination
   ____ extra
   ____ enough

9. ____ stop
   ____ step
   ____ stopped
   ____ stepped

10. ____ could
    ____ cannot
    ____ cold
    ____ cramp

11. ____ far
    ____ farm
    ____ firm
    ____ father

12. ____ tough
    ____ though
    ____ thought
    ____ thorough

STOP !

II. Put an "X" by the answer that tells what the underlined word means.

Sample:

friend means:

__x_pal

____day

____enemy

____cooked

1. quickly means:

____slow

____fast

____quiet

____quit

2. to sew means:

____to cut a piece of wood

____to see someone

____to use a needle and thread

____to plant seeds

3. lamb means:

____a light

____a part of a tree

____a baby sheep

____to come down

4. exist means:

____to die

____to go out

____to live

____to eat

5. between means

____above

____below

____next

____in the middle

6. gem means:

____a boy

____a place today

____a jewel

____a thing to chew

7. <u>struck</u> means:

____stick

____hit

____bit

____truck

STOP!

IV. Read the story and put an "X" by the correct answer for each question.

Maria and Carlos went to the Bronx zoo on Saturday. They bought two bags of peanuts. When Maria and Carlos got hungry, they ate one bag of peanuts. They saved the other bag of peanuts to feed the elephants.

SAMPLE: What is the name of the girl in this story?

____Susan

_X__Maria

____Angela

____Juanita

1. What did Maria and Carlos go to the zoo?

____on Sunday

____on Monday

____on Saturday

____on Thursday

2. What did Maria and Carlos buy in the zoo?

____no bags of peanuts

____ten bags of peanuts

____three bags of peanuts

____two bags of peanuts

3. The children ate one bag of peanuts:

____when they awoke from their nap

____when they met their friend

____the next day

____when they got hungry

IV. Johnny came down the steps onto the stoop. He sat with his arms resting on his knees and his chin on his hands. Johnny blew large pink bubbles through his lips. Often he would twist his left sneaker to look at the big holes on the bottom.

What time could it be now? Mom had said one hour, but to Johnny, it seemed that he'd be waiting for three. The other boys in the block had already gone down to the corner with their sticks. Mom had not yet come back, and soon the game would start.

Johnny began to worry. What could they do? Mom had been serious when she'd said, "No ball playing until you get new sneakers." So, he didn't dare to go. But soon, the bigger boys would come and would not let Johnny play. "What to do, what to do."

Just then, Johnny looked sadly towards the corner. Suddenly, Johnny saw mom coming, with many bags in her arms. "Yea," yelled Johnny as he jumped over two garbage cans and a fire pump on his wild run up the street.

4. What did Johnny do while he waited for mom?

____blow bubbles

____take off his sneakers

____play ball

5. What kind of ball were the boys playing?

____football

____basketball

____stickball

6. Why did Johnny wait?

____because mom told him to

____he was afraid of the bigger boys

____he had bubble gum on his sneakers

7. What did Johnny want most?

____some more bubble gum

____to play in the ball game

____to wait for mom

(IV.)

About forty years ago, a black girl was born in the big city of Chicago. Her name was Lorraine Hansbury.

When she was a little girl, Lorraine spent most of her summers out of doors. In the day, she played street games with her friends. One game was called "giant steps."

On hot summer nights, the family would often go to the park where Mr. and Mrs. Hansbury would tell the children stories about their own childhood down South.

As she grew up, Lorraine went to school where she learned to read and to write poems and plays. She attended the university and then came to New York to continue her education and playwriting.

This black writer won many honors for her play called "A Raisin in the Sun." In this play, Lorraine writes about the life of a black family, like hers, living in Chicago. This play is also a movie. One day you may see it on television. If so, you might remember that it was written by a young black woman.

When you are older, you may want to read books about her life and works. Maybe one day, you might even write poems or a play about your life.

8. "Giant steps" is a game that

____city children play

____grown-ups play

____giants play

9. Because she remembered many stories and she learned to read, Lorraine could

____not go to school

____write her own poems and plays

____not become famous

10. "A Raisin in the Sun" is a play for which Lorraine

____was scolded

____was angry about

____won many honors

11. "A Raisin in the Sun" can be seen in the movies and

____at the store

____on television

____at the farm

12. This story is about

____a black educator

____a black writer

____a black game

STOP!

V. Mark an "X" by the word that <u>rhymes</u> with the <u>underlined</u> word.

SAMPLE: <u>end</u>

____sent

____sand

__x__spend

1. <u>milk</u>

____skill

____silk

____sick

2. <u>thigh</u>

____they

____there

____high

3. <u>fast</u>

____fuss

____fist

____mast

4. <u>struck</u>

____strike

____truck

____track

5. <u>ship</u>

____shop

____chip

____shut

6. <u>talk</u>

____silk

____take

____walk

7. <u>call</u>

____roll

____small

____full

8. <u>make</u>

____made

____mate

____cake

9. <u>shoe</u>

____shell

____tea

____too

10. <u>suit</u>

____sweet

____ sued

____fruit

11. <u>box</u>

____ax

____max

____rocks

Appendix XI

Administration Instructions

Reading Mini-Test I*

Test III: Part A: Sentences (Time: 3 minutes)

To administer Test 3: Part A: Sentences, SAY TO THE PUPILS:

> Open your book, like this, to page 3. Fold your book so that only page 3 is showing.

Demonstrate. See that pupils have only page [3] facing up. THEN SAY:

> There are some pictures and some sentences on this page. Find the picture of the toys at the top of the page. Put your finger on the picture.

Check to see that all children have the place.

> See the sentences next to the picture.

Point to your own booklet, or reproduce the sample on the board and refer to it. THEN SAY:

> Let's read the sentences together. They are:
> "The toys are on the table."
> "These toys are for children."
> "The top is bigger than the ball."
> Which sentence tells about the picture?

*****

> Yes, "These toys are for children," tells about the picture. You see that the little space in front of the sentence, "These toys are for children," has been filled in.
> On this page, there are more pictures and sentences. Look at each picture; then read each of the three sentences beside the picture carefully. You are to fill in the little space in front of the sentence that tells about the picture. Do all the pictures on this page. Keep working until you come to the bottom of the page where it says STOP. Then put your pencil down and wait. Do you all know what to do?

*Test III administration instructions reproduced from Metropolitan Achievement Test, Teacher's Directions, copyright © 1970 by Harcourt Brace Jovanovich, Inc. Reproduced by special permission of the publisher.

*****

After 3 minutes, SAY:

Stop! Put your pencil down.

<u>Test III: Park B: Stories</u> (Time: 2 minutes)

THEN SAY:

Now turn your book to page <u> 4 </u>. Fold your book so that only page <u> 4 </u> is showing.

*****

Demonstrate. Make sure that everyone has the correct page. THEN SAY:

There are some stories on this page. This is what you are to do. First you read the story. Think what the story is about. Then you read the words under the story. When you find the work that tells what the story is about, fill in the space in front of that word. Let's read the first story together. It says:

"I am like a short coat.
I keep you warm.
I am a ____."

Now let's read the words underneath the story. They are -- "jacket, raincoat, dress." Which of these tells what the story is about?

Encourage reply.

Yes, "jacket." The space in front of "jacket" has been filled in. Now do you know how to do the stories? You are going to do both the stories on this page by yourself. Work across the page like this.

Demonstrate on page <u> 4 </u> of your own booklet.

Read the first story. Think what it is about. Find the word (or words) that tells what the story is about. Fill in the space in front of the word (or words) with your pencil. When you finish the first story go on to the second story, which is right beside the first story.

Point to your own booklet.

When you come to where it says STOP at the bottom of the page, put your pencil down and wait.

Ready. Go!

After 2 minutes, SAY:

Stop!

THEN SAY:

Turn to the next page.

Demonstrate.

<u>Test III: Part B: Stories (CONT.)</u> (Time: 5 minutes)

THEN SAY:

Here are some more stories for you to read. Find the story at the top of the page beginning with the words "Ted has a puppy." Put your finger on these words. Let's read this story together. It says:

"Ted has a puppy.
His name is Happy.
Ted and Happy like to play."

Now look at the questions next to the story. The first question says: "The pet is a -- dog, boy, toy." What does the story tell us that the pet is?

Encourage reply.

Yes, the story tells us that the pet is a dog, so the space in front of the word "dog" has been filled in. The next question says "Ted and Happy like to -- play, work, look." What is the answer?

Encourage reply.

Yes, the answer is "play," so fill in the space in front of the word "play."

Read each story on this page to yourself. Mark the answer to the questions in the same way as they were marked in the story we just read.

Show page <u>5</u> .

Keep working until you come to where it says STOP. Then put your pencil down and wait. Are there any questions about what you are to do?

*****

Ready. Go!

*****

Give no further help. Check to see that pupils turn to page [5] and do not go on to Test IV.

*****

After 5 minutes, SAY:

Stop! Put your pencil down.

Appendix XII

Administration Instructions

Reading Mini-Test II*

Test III: Part A: Sentences (Time: 2 minutes)

When starting Test 3: Part A: Sentences, SAY TO THE PUPILS:

> Open your book like this to page 3 .

Demonstrate. See that pupils have page 3 folded back.

THEN SAY:

> There are some pictures and some sentences on this page. Find the picture of the toys at the top of the page. Put your finger on the picture.

Check to see that all children have the place.

> See the sentences next to the picture.

Point to your own booklet, or reproduce the sample on the board and refer to it. THEN SAY:

> Let's read the sentences together. They are:
> "These toys are for children."
> "These toys are on the table."
> "The top is bigger than the ball."
> Which sentence tells about the picture?

*****

> Yes, "These toys are for children" tells about the picture. You see that the little space in front of the sentence, "These toys are for children," has been filled in. On this page and the next, there are more pictures and sentences. Look at each picture; then read each of the three sentences beside the picture carefully. You are to fill in the space in front of the sentence that tells about the picture.

*****

> Do all the pictures on this page.

*Test III administration instructions reproduced from Metropolitan Achievement Test, Teacher's Directions, copyright © 1970 by Harcourt Brace Jovanovich, Inc. Reproduced by special permission of the publisher.

*****

Keep working until you come to where it says STOP. Then put your pencil down and wait. Do you all know what to do?

*****

Ready. Go!

*****

At the end of 2 minutes, SAY:

STOP! Put your pencil down.

<u>Test III: Part B: Stories</u> (Time: 6 minutes)

To administer Test 3: Part B, Stories, SAY:

Open your book to page <u>4</u>. Fold the page back like this so that only page <u>4</u> is showing.

*****

There are some stories on this page. Find the story at the top of the page beginning with the words "Mother made a cake." Put your finger on these words. Let's read this story together. It says:

"Mother made a cake.
She put candles on it.
The candles told how old Jane was."

Now look for the questions in the same box with the story. The first question says:
"Mother was getting ready for -- a birthday, summer, Thanksgiving."
What does the story tell us Mother was getting ready for?

Encourage reply.

Yes, the story tells us that Mother was getting ready for "a birthday," so the space in front of the words "a birhtday" has been filled in. The next question says: "On the cake, Mother put -- candles, candy,ice cream." Fill in the correct answer.

*****

The answer is "candles," so the space in front of the word "candles" should be filled in. Read each story on this page to yourself. Fill in the space in front of the correct answers to the questions that go with each story.

Keep working until you come to where it says STOP at the bottom of the page. Then put your pencil down and wait.

Are there any questions about what you are to do?

*****

Start with the story just below the sample story.

Point to the first story in your booklet. THEN SAY:

Ready. Go!

*****

At the end of 6 minutes, SAY:

STOP!

Now put your pencil down.

Appendix XIII

Administration Instructions - Teacher-Made Reading Test with Mini-Test I*

Be sure that each pupil has a test booklet corresponding to the Metropolitan Achievement Test level which he took recently - Primary I or Primary II.

Ask each pupil to write his name and class (2-1, 2-2, etc.) on the front of his test booklet.

* * * * * * *

Test I (Time: about 8 minutes)

Say to the pupils:

Turn to page 1. For each question on this page, I will read one word. Listen carefully to the word I say, and mark an "X" by that word.

Look at the sample question at the top of the page. I will say one of the words -- sing. See that an "X" has been marked next to the word "sing".

Ask if there are any questions, and clarify the instructions if necessary.

Then say to the pupils:

Now begin with question 1 and mark the word that I say with an "X".

Read the following list:

1.	sheep	7.	much
2.	ship	8.	enough
3.	send	9.	step
4.	high	10.	cold
5.	horse	11.	farm
6.	child	12.	though

Say to the pupils:

Stop. Put your pencils down and turn to page 2.

* * * * * * *

*Test III administration instructions reproduced from Metropolitan Achievement Test, Teacher's Directions, copyright © 1970 by Harcourt Brace Jovanovich, Inc. Reproduced by special permission of the publisher.

Test II. (Time: about 5 minutes)

Say to the pupils:

For each question on this page, put an "X" by the answer that tells what the underlined word means.

Look at the sample question at the top of the page. It says, "friend means: pal, day, enemy, cooked." An "X" has been put next to the answer - pal.

Now go ahead and read each of the questions on this page and put an "X" by the right answer. Stop at the end of the page and put your pencils down.

Clarify the instructions if this seems necessary.

Test III: Part A: Sentences (Time: 3 minutes)

To administer Test 3: Part A: Sentences, SAY TO THE PUPILS:

Open your book, like this, to page 3. Fold your book so that page 3 is showing.

Demonstrate. See that pupils have only page 3 facing up. THEN SAY:

Let's read the sentences together. They are:
"The toys are on the table."
"These toys are for children."
"The top is bigger than the ball."
Which sentence tells about the picture?

*****

Yes, "These toys are for children," tells about the picture. You see that the little space in front of the sentence, "These toys are for children," has been filled in.
On this page, there are more pictures and sentences. Look at each picture; then read each of the three sentences beside the picture carefully. You are to fill in the little space in front of the sentence that tells about the picture. Do all the pictures on this page. Keep working until you come to the bottom of the page where it says STOP. Then put your pencil down and wait. Do you all know what to do?

*****

Ready. Go!

Give no further help. Check to see that pupils do not go on to the next page.

*****

After 3 minutes, SAY:

> Stop! Put your pencil down.

Test III: Part B: Stories (Time: 2 minutes)

THEN SAY:

> Now turn your book to page 4. Fold your book so that only page 4 is showing.

*****

Demonstrate. Make sure that everyone has the correct page. THEN SAY:

> There are some stories on this page. This is what you are to do. First you read the story. Think what the story is about. Then you read the words under the story. When you find the word that tells what the story is about, fill in the space in front of that word. Let's read the first story together. It says:
> "I am like a short coat.
> I keep you warm.
> I am a ____."
> Now let's read the words underneath the story. They are -- "jacket, raincoat, dress." Which of these tells what the story is about?

Encourage reply.

> Yes, "jacket." The space in front of "jacket" has been filled in. Now do you know how to do the stories? You are going to do both the stories on this page by yourself. Work across the page like this.

Demonstrate on page 4 of your own booklet.

> Read the first story. Think what it is about. Find the word (or words) that tells what the story is about. Fill in the space in front of the word (or words) with your pencil. When you finish the first story go on to the second story, which is right beside the first story.

Point to your own booklet.

> When you come to where it says STOP at the bottom of the page, put your pencil down and wait.

Ready. Go!

After 2 minutes, SAY:

Stop!

THEN SAY:

Turn to the next page.

Demonstrate.

<u>Test III: Part B: Stories (CONT.)</u> (Time: 5 minutes)

THEN SAY:

Here are some more stories for you to read. Find the story at the top of the page beginning with the words "Ted has a puppy." Put your finger on these words. Let's read this story together. It says:

"Ted has a puppy.
His name is Happy.
Ted and Happy like to play."

Now look at the questions next to the story. The first question says: "The pet is a -- dog, boy, toy." What does the story tell us that the pet is?

Encourage reply.

Yes, the story tells us that the pet is a dog, so the space in front of the word "dog" has been filled in. The next question says "Ted and Happy like to -- play, work, look." What is the answer?

Encourage reply.

Yes, the answer is "play," so fill in the space in front of the word "play."

Read each story on this page to yourself. Mark the answers to the questions in the same way as they were marked in the story we just read.

Show page <u>5</u>.

Keep working until you come to where it says STOP. Then put your pencil down and wait. Are there any questions about what you are to do?

*****

Ready. Go!

*****

Give no further help. Check to see that pupils turn to page 5 and do not go on to Test IV.

*****

After 5 minutes, SAY:

Stop! Put your pencil down.

* * * * * * * * * * *

If your pupils are tired or restless, give them a rest period for a few minutes before beginning Test IV.

Test IV (Time: about 15 minutes)

Ask the pupils to turn to section IV of their booklets (there is no page number for this section), and be sure that they have the correct page.

Say to the pupils:

You now have some more stories to read. After you read each story, there will be some questions to answer about the story. Put an "X" by the right answer for each question.

Look at the sample question right under the first story. It says, "What is the name of the girl in this story? -- Susan, Maria, Angela, Juanita." You see that an "X" has been put by the right answer, which is Maria.

Now go ahead and read the stories on this page and on the next 3 pages. Stop when you have answered question 12, and put your pencils down.

Clarify the instructions if necessary.

* * * * * * *

Test V (Time: about 10 minutes)

Ask the pupils to turn to section V (no page number), and check to see that they have the correct page.

Say to the pupils:

For each question on this page, mark an "X" by the word that rhymes with the underlined word.

Look at the sample question at the top of the page. The word that is underlined is end. An "X" has been marked by the word that rhymes with end, which is "spend."

Now go ahead and do each of the questions on this page. When you have finished put your pencils down.

Clarify the instructions if necessary.

* * * * * * *

Thank you!

Appendix XIV

Administration Instructions - Teacher-Made Reading Test with Mini-Test II*

Be sure that each pupil has a test booklet corresponding to the Metropolitan Achievement Test level which he took recently - Primary I or Primary II.

Ask each pupil to write his name and class (2-1, 2-2, etc.) on the front of his test booklet.

* * * * * * *

Test I (Time: about 8 minutes)

Say to the pupils:

Turn to page 1. For each question on this page, I will read one word. Listen carefully to the word I say, and mark an "X" by that word.

Look at the sample question at the top of the page. I will say one of the words -- sing. See that an "X" has been marked next to the word "sing".

Ask if there are any questions, and clarify the instructions if necessary.

Then say to the pupils:

Now begin with question 1 and mark the word that I say with an "X".

Read the following list:

1. sheep	7. much
2. ship	8. enough
3. send	9. step
4. high	10. cold
5. horse	11. farm
6. child	12. though

Say to the pupils:

Stop. Put your pencils down and turn to page 2.

* * * * * * *

Test II. (Time: about 5 minutes)

*Test III administration instructions reproduced from Metropolitan Achievement Test, Teacher's Directions, copyright © 1970 by Harcourt Brace Jovanovich, Inc. Reproduced by special permission of the publisher.

Say to the pupils:

For each question on this page, put an "X" by the answer that tells what the underlined word means.

Look at the sample question at the top of the page. It says, "friend means: pal, day, enemy, cooked," An "X" has been put next to the answer - pal.

Now go ahead and read each of the questions on this page and put an "X" by the right answer. Stop at the end of the page and put your pencils down.

Clarify the instructions if this seems necessary.

Test III: Part A: Sentences (Time: 2 minutes)

When starting Test 3: Part A: Sentences, SAY TO THE PUPILS:

Open your book like this to page 3.

Demonstrate. See that pupils have page 3 folded back.

THEN SAY:

There are some pictures and some sentences on this page. Find the picture of the toys at the top of the page. Put your finger on the top of the page. Put your finger on the picture.

Check to see that all children have the place.

See the sentences next to the picture.

Point to your own booklet, or reproduce the sample on the board and refer to it. THEN SAY:

Let's read the sentences together. They are:
"These toys are for children."
"These toys are on the table."
"The top is bigger than the ball."
Which sentence tells about the picture?

*****

Yes, "These toys are for children" tells about the picture. You see that the little space in front of the sentence, "These toys are for children," has been filled in . On this page and the next, there are more pictures and sentences. Look at each picture; then read each of the three sentences beside the picture carefully. You are to fill in the space in front of the sentence that tells about the picture.

*****

Do all the pictures on this page.

*****

Keep working until you come to where it says STOP. Then put your pencil down and wait. Do you all know what to do?

*****

Ready. Go!

*****

At the end of 2 minutes, SAY:

STOP! Put your pencil down.

Test III: Part B: Stories (Time: 6 minutes)

To administer Test 3: Part B, Stories, SAY:

Open your book to page 4. Fold the page back like this so that only page 4 is showing.

*****

There are some stories on this page. Find the story at the top of the page beginning with the words "Mother made a cake." Put your finger on these words. Let's read this story together. It says:

"Mother made a cake.
She put candles on it.
The candles told how old Jane was."

Now look for the questions in the same box with the story. The first question says:
"Mother was getting ready for -- a birthday, summer, Thanksgiving." What does the story tell us Mother was getting ready for?

Encourage reply.

Yes, the story tells us that Mother was getting ready for "a birthday" has been filled in. The next question says: "On the cake, Mother put -- candles, candy, ice cream." Fill in the correct answer.

*****

The answer is "candles," so the space in front of the word "candles" should be filled in. Read each story on this page to yourself. Fill in the space in front of the correct answers to the questions that go with each story.

Keep working until you come to where it says STOP at the bottom of the page. Then put your pencil down and wait.

Are there any questions about what you are to do?

*****

Start with the story just below the sample story.

Point to the first story in your booklet. THEN SAY:

Ready. Go!

*****

At the end of 6 minutes, SAY:

STOP!

Now put your pencil down.

If your pupils are tired or restless, give them a rest period for a few minutes before beginning Test IV.

Test IV (Time: about 15 minutes)

Ask the pupils to turn to section IV of their booklets (there is no page number for this section), and be sure that they have the correct page.

Say to the pupils:

You now have some more stories to read. After you read each story, there will be some questions to answer about the story. Put an "X" by the right answer for each question.

Look at the sample question right under the first story. It says, "What is the name of the girl in this story? -- Susan, Maria, Angela, Juanita." You see that an "X" has been put by the right answer, which is Maria.

Now go ahead and read the stories on this page and on the next 3 pages. Stop when you have answered question 12, and put your pencils down.

Clarify the instructions if necessary.

* * * * * * *

Test V (Time: about 10 minutes)

Ask the pupils to turn to section V (no page number), and check to see that they have the correct page.

Say to the pupils:

For each question on this page, mark an "X" by the word that rhymes with the underlined word.

Look at the sample question at the top of the page. The word that is underlined is end. An "X" has been marked by the word that rhymes with end, which is "spend".

Now go ahead and do each of the questions on this page. When you have finished put your pencils down.

Clarify the instructions if necessary.

* * * * * *

Thank you!

Appendix XV

Estimated Means and Standard Deviations for 11 Stratified Random Samples (N=153)

Sample	Actual and Estimated Means	Actual and Estimated Standard Deviations	$\mu-\hat{\mu}$
Complete Test	45.32	11.65	0
Post hoc mini-test	44.18	10.90	1.18
2	44.26	11.74	1.06
3	45.05	12.41	.27
4	44.10	12.20	1.22
5	47.60	10.33	-2.28
6	45.05	12.12	.27
7	45.01	11.24	.31
8	44.45	12.47	.87
9	46.62	10.78	-1.30
10	47.88	10.58	-2.56
11	46.23	11.46	- .91

Appendix XVI

Estimated Means and Standard Deviations
for 10 Random Samples
Drawn from Complete Test (N=153)

Sample	Actual and Estimated Means	Actual and Estimated Standard Deviations	$\mu - \hat{\mu}$
Complete Test	45.32	11.65	0
1	44.34	12.62	.98
2	46.19	10.57	- .87
3	46.78	10.40	-1.46
4	46.23	11.16	- .91
5	47.05	9.87	-1.73
6	45.95	11.25	- .63
7	42.92	12.28	2.40
8	44.34	9.88	.98
9	48.66*	9.18	3.54
10	46.78	10.82	1.46

*Estimated mean not within plus or minus one SEM.

Appendix XVII

Estimated Means for 100 Random Samples (N=153)

Sample	$\hat{\mu}$	Sample	$\hat{\mu}$	Sample	$\hat{\mu}$	Sample	$\hat{\mu}$
1	44.29	26	45.39	51	47.08	76	45.60
2	44.99	27	46.53	52	45.90	77	43.46
3	44.05	28	48.93	53	44.17	78	47.91
4	47.60	29	46.06	54	40.94	79	47.16
5	45.03	30	44.68	55	43.02	80	45.82
6	44.99	31	44.76	56	47.00	81	45.07
7	44.29	32	43.14	57	48.58	82	43.89
8	46.57	33	45.51	58	44.80	83	45.27
9	47.91	34	45.74	59	43.58	84	48.54
10	46.18	35	42.99	60	44.64	85	47.04
11	45.47	36	45.82	61	45.86	86	45.86
12	44.21	37	45.39	62	43.02	87	44.44
13	47.56	38	44.68	63	44.96	88	45.82
14	10.93	39	48.93	64	41.17	89	45.55
15	51.61	40	46.06	65	46.26	90	46.73
16	47.60	41	44.76	66	45.31	91	45.43
17	42.63	42	48.23	67	48.74	92	45.27
18	42.95	43	44.92	68	40.86	93	45.23
19	45.70	44	43.22	69	40.46	94	42.95
20	49.72	45	44.68	70	47.00	95	43.42
21	43.89	46	45.39	71	45.47	96	44.21
22	46.22	47	42.08	72	45.31	97	42.08
23	46.96	48	45.98	73	46.33	98	44.75
24	43.22	49	44.33	74	43.50	99	44.05
25	46.77	50	47.48	75	47.60	100	42.81

Appendix XVIII

Estimated Standard Deviations for 100 Random Samples (N=153)

Sample	$\hat{\sigma}$	Sample	$\hat{\sigma}$	Sample	$\hat{\sigma}$	Sample	$\hat{\sigma}$
1	12.02	26	11.46	51	10.85	76	11.62
2	12.58	27	11.17	52	11.72	77	12.23
3	12.47	28	10.15	53	12.42	78	10.94
4	10.47	29	11.21	54	13.57	79	10.05
5	12.37	30	11.83	55	11.82	80	11.13
6	11.46	31	11.60	56	11.08	81	10.42
7	12.44	32	11.83	57	10.51	82	11.52
8	10.96	33	11.23	58	10.94	83	11.31
9	10.72	34	10.66	59	12.15	84	10.66
10	11.59	35	11.26	60	12.74	85	11.02
11	10.91	36	11.62	61	11.62	86	11.62
12	12.14	37	11.46	62	11.82	87	12.08
13	11.00	38	10.90	63	11.51	88	11.59
14	10.93	39	10.15	64	12.50	89	12.06
15	12.32	40	10.54	65	11.75	90	10.86
16	10.88	41	11.60	66	12.47	91	11.94
17	12.82	42	11.53	67	10.67	92	11.87
18	12.08	43	11.54	68	13.63	93	12.13
19	12.81	44	12.44	69	12.60	94	12.13
20	10.76	45	11.83	70	11.08	95	11.74
21	10.07	46	11.46	71	11.82	96	11.98
22	10.56	47	13.46	72	12.47	97	12.74
23	10.89	48	10.83	73	11.62	98	11.26
24	11.88	49	11.98	74	12.09	99	12.21
25	10.58	50	11.94	75	10.37	100	11.99

Appendix XIX

Estimated KR21 Coefficients for 100 Random Samples (N=153)

Sample	$\hat{KR21}$	Sample	$\hat{KR21}$	Sample	$\hat{KR21}$	Sample	$\hat{KR21}$
1	.92	26	.92	51	.91	76	.92
2	.93	27	.92	52	.92	77	.92
3	.93	28	.91	53	.93	78	.92
4	.91	29	.92	54	.94	79	.90
5	.93	30	.92	55	.92	80	.91
6	.92	31	.92	56	.92	81	.90
7	.93	32	.92	57	.91	82	.91
8	.91	33	.91	58	.91	83	.91
9	.91	34	.90	59	.92	84	.92
10	.92	35	.91	60	.93	85	.92
11	.91	36	.92	61	.92	86	.92
12	.93	37	.92	62	.92	87	.92
13	.92	38	.91	63	.92	88	.92
14	.94	39	.91	64	.92	89	.93
15	.95	40	.90	65	.93	90	.91
16	.92	41	.92	66	.93	91	.93
17	.93	42	.93	67	.92	92	.92
18	.92	43	.92	68	.94	93	.93
19	.94	44	.93	69	.92	94	.93
20	.92	45	.92	70	.92	95	.92
21	.88	46	.92	71	.92	96	.92
22	.90	47	.94	72	.93	97	.93
23	.91	48	.91	73	.92	98	.91
24	.92	49	.92	74	.92	99	.93
25	.91	50	.93	75	.91	100	.92

Appendix XX

Actual KR21 Coefficients for 100 Random Samples (N=153)

Sample	KR21	Sample	KR21	Sample	KR21	Sample	KR21
1	.79	26	.77	51	.75	76	.77
2	.79	27	.76	52	.79	77	.77
3	.80	28	.74	53	.78	78	.76
4	.74	29	.76	54	.81	79	.72
5	.79	30	.77	55	.76	80	.76
6	.76	31	.76	56	.76	81	.72
7	.80	32	.77	57	.76	82	.75
8	.75	33	.76	58	.74	83	.76
9	.75	34	.74	59	.78	84	.76
10	.78	35	.74	60	.80	85	.76
11	.75	36	.78	61	.77	86	.77
12	.78	37	.77	62	.76	87	.78
13	.75	38	.74	63	.76	88	.77
14	.77	39	.74	64	.79	89	.79
15	.80	40	.73	65	.78	90	.75
16	.75	41	.76	66	.79	91	.78
17	.80	42	.78	67	.75	92	.78
18	.77	43	.76	68	.81	93	.79
19	.80	44	.78	69	.78	94	.79
20	.76	45	.77	70	.76	95	.76
21	.74	46	.77	71	.77	96	.78
22	.73	47	.81	72	.79	97	.79
23	.75	48	.74	73	.78	98	.77
24	.76	49	.78	74	.77	99	.79
25	.74	50	.79	75	.74	100	.77

Appendix XXI

Copyright Letter (I)

Harcourt Brace Jovanovich, Inc.

757 Third Avenue, New York, N.Y. 10017 Telephone: 572-5000

May 19, 1972

Miss Helen Spilman
Division of Teacher Education
The City University of New York
535 East Eightieth Street
New York, N.Y.10021

Dear Miss Spilman:

This is in reply to your letter of May 5*to Dr. Bligh in which you listed the items that you wish to reproduce.

We are happy to grant you permission to reproduce the items stated in your letter of May 5 provided you use the credit line shown below.

Thank you for your interest and cooperation. If I can be of further assistance, please do not hesitate to contact me.

Sincerely yours,

Collette Mullen

Collette Mullen
Administrative Assistant

cc: Rita Vaughn
Joan Sonnenschein

Reproduced from Metropolitan Achievement Test,
copyright © 1970 by Harcourt Brace Jovanovich, Inc.
Reproduced by special permission of the publisher.
or
Reproduced from Stanford Achievement Test,
copyright © 1964-1966 by Harcourt Brace Jovanovich, Inc.
Reproduced by special permission of the publisher.

*See Appendix XXII

Appendix XXII

Copyright Letter (II)

OFFICE OF TEACHER EDUCATION
THE CITY UNIVERSITY OF NEW YORK
1411 BROADWAY
NEW YORK, NEW YORK 10018
212-239-7430

May 5, 1972

Dr. Harold Bligh
Harcourt Brace Jovanovich Publications
757 Third Avenue
New York, New York 10017

Dear Dr. Bligh:

This is the information which you requested in our telephone conversation earlier this week regarding my request to duplicate some item samples from tests for which your company holds the copyright. Thank you for your advice on the format to use for protecting your copyright.

The items which I will be reporducing are:

1. Metropolitan Achievement Tests, Primary I, Form F, Test 3: Reading: items 6, 7, 12, 15, 17, 25, 27, 28, 39, 41, 42 (200 copies)

2. Metropolitan Achievement Tests, Primary II, Form F, Test 3: Reading: items 4, 7, 8, 20, 21, 22, 24, 36, 38, 39, 40 (70 copies)

3. Stanford Achievement Test, Primary I, Forms Y and X, Test 6: Arithmetic: items 2, 6, 8, 17, 26, 28, 30, 31, 34, 38, 39, 45, 49, 52, 54, 63 (100 copies for each form)

Further information on how the item samples will be used is included in my letter of April 3, 1972.*

Thank you again for your assistance. I hope to speak with you personally regarding the availability of item data for these tests.

Sincerely,

Helen Spilman

Helen Spilman

*See Appendix XXIII

Appendix XXIII

Copyright Letter (III)

OFFICE OF TEACHER EDUCATION
THE CITY UNIVERSITY OF NEW YORK
1411 BROADWAY
NEW YORK, NEW YORK 10018
212-239-7430

April 3, 1972

Dr. Harold Bligh
Harcourt Brace Jovanovich Publications
757 Third Avenue
New York, New York 10017

Dear Dr. Bligh:

I appreciate having had the opportunity to discuss with you some aspects of my research project this past Friday. As I explained, I will need item data (item difficulties, item-total test correlations, etc.) for two tests published by Harcourt Brace Jovanovich.

This item data will be used in determining criteria for the selection of item samples. The purpose of the sampling procedure in this case is to determine if single item samples can provide "good" estimates of group means and standard deviations.

The tests for which the item data are required are:

1. Metropolitan Achievement Tests, Primary I, Reading Tests 1, 2, & 3 (1970), for Form H and a parallel form (F or G)

2. Metropolitan Achievement Tests, Primary II, Reading Tests 1 through 4 (1970), for Form H and a parallel form (F or G)

3. Stanford Achievement Test, Primary I, Arithmetic Computation, Test 7 (1964), for Form W and a parallel form (X or Y)

I would like to have your permission to administer samples of items from these tests to approximately 300 second grade pupils in two elementary schools in the Bronx. One half of the pupils will be administered only the item sample. The other half will have the item sample administered within a teacher-made test. Under both administration conditions the source of the standardized test items will be identified on the test forms. We plan to administer the item samples during May of this year.

Thank you very much for your interest and your assistance. I look forward to meeting you in the near future.

Sincerely,

Helen Spilman

Helen Spilman

cc: Dr. Max Weiner

REFERENCES

Barcikowski, R. S. A monte carlo study of item sampling (versus traditional sampling) for norm construction. Journal of Educational Measurement, 1972, 9, 209-214.

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

Cook, D. L. and Stufflebeam, D. L. Estimating norms from variable size item and examinee samples. Journal of Educational Psychology, 1967, 4, 27-33.

Feldt, L. S. A note on the use of confidence bands to evaluate the reliability of a difference between two scores. American Educational Research Journal, 1967, 4, 139-145.

Gulliksen, H. Theory of Mental Tests. New York: John Wiley and Sons, Inc., 1950.

Harcourt Brace Jovanovich, Inc. Stanford Achievement Test. New York, 1964-1966.

Harcourt Brace Jovanovich, Inc. Metropolitan Achievement Test. New York, 1970.

Hays, W. L. Statistics. New York: Holt, Rinehart and Winston, 1963.

Husek, T. R. and Sirotnik, K. Matrix sampling in educational research: An empirical investigation. Paper presented at the Convention of the American Educational Research Association, 1968.

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

Johnson, M. C. and Lord, F. M. An empirical study of the stability of a group mean in relation to the distribution of test items among students. Educational and Psychological Measurement, 1958, 18, 325-329.

Kleinke, D. J. A linear prediction approach to developing test norms based on matrix-sampling. Educational and Psychological Measurement, 1972, 32, 75-84.

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

Lord, F. M. and Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley Publishing Co., 1968.

Mehrens, W. A. and Lehmann, I. J. Standardized Tests in Education. New York: Holt, Rinehart and Winston, Inc., 1969.

Owens, T. R. and 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 norms and standard deviations from partial data: An empirical check on Lord's item sampling technique. Educational and Psychological Measurement, 1964, 24, 623-630.

Pugh, R. C. Empirical evidence on the application of Lord's sampling technique to Likert items. Journal of Experimental Education. 1971, 39, 54-56.

Sax, G. S. and Cromack, T. R. The effects of various forms of item arrangements on test performance. Journal of Educational Measurement, 1966, 3, 309-311.

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

Shoemaker, D. M. Item-examinee sampling procedures and associated standard errors in estimating test parameters. Journal of Educational Measurement, 1970(b), 7, 255-262.

Shoemaker, D. M. Test statistics as a function of item arrangement. Journal of Experimental Education, 1970(c), 39, 85-88.

Shoemaker, D. M. Further results on the standard errors of estimate associated with item-examinee sampling procedures. Journal of Educational Measurement, 1971(a), 8, 215-220.

Shoemaker, D. M. An application of item-examinee sampling to scaling attitudes. Journal of Educational Measurement, 1971(b), 8, 279-282.

Shoemaker, D. M. Principles and procedures of multiple matrix sampling. Southwest Regional Laboratory for Educational Research and Development, Technical Report 34, 1971(c).

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