DOCUMENT RESUME

ED 064 342	TM 001 507
AUTHOR	Montague, Margariete A.
TITLE	Use of Matrix Sampling Procedures to Assess Achievement in Solving Open Addition and Subtraction Sentences.
PUB DATE	Apr 72
NOTE	16p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Analysis of Variance; Elementary School Students; Group Tests; *Item Analysis; *Sampling; *Test Construction; *Test Results

ABSTRACT

This study investigated the feasibility of concurrently and randomly sampling examinees and items in order to estimate group achievement. Seven 32-item tests reflecting a 640-item universe of simple open sentences were used such that item selection (random, systematic) and assignment (random, systematic) of items (four, eight, sixteen) to forms were varied. Twenty-four second or third grade populations were randomly selected. Analysis of Variance was used to examine the data. Nonsignificant differences were observed with respect to item selection, item assignment, and number of items per form. Results support the appropriateness of the procedure for estimating group achievement. (Author)

USE OF MATRIX SAMPLING PROCEDURES TO ASSESS ACHIEVEMENT IN SOLVING OPEN ADDITION AND SUBTRACTION SENTENCES

Margariete A. Montague
University of Wyoming
Laramie, Wyoming

A paper presented at the annual meeting of the
American Educational Research Association
April 3-7, Chicago, Illinois

USE OF MATRIX SAMPLING PROCEDURES TO ASSESS ACHIEVEMENT IN SOLVING OPEN ADDITION AND SUBTRACTION SENTENCES

Margariete A. Montague
The University of Wyoming
Laramie, Wyoming 82070

ABSTRACT

This study investigated the feasibility of concurrently and randomly sampling examinees and items in order to estimate group achievement.

Seven 32-item tests reflecting a 640-item universe of simple open sentences were used such that item selection (random, systematic) and assignment (random, systematic) of items (four, eight, sixteen) to forms were varied. Twenty-four second or third grade populations were randomly selected. Analysis of Variance was used to examine the data.

Nonsignificant differences were observed with respect to item selection, item assignment, and number of items per form. Results support the appropriateness of the procedure for estimating group achievement.

OBJECTIVES

The problem investigated was the feasibility of concurrently and randomly sampling examinees and items in order to obtain group data generalizable to a universe of examinees and to a universe of items. More specifically, the purpose of the study was to use matrix sampling to evaluate the ability of randomly selected populations of second and third grade children to solve simple equations derived from $a \pm b = c$ and from $c = a \pm b$.

The term "matrix sampling" is used here to denote the concurrent sampling of M examinees from a universe of $\mathcal{M}$ examinees and N items from a universe of $\mathcal{N}$ mathematical items. The M examinees were randomly partitioned into k mutually exclusive samples of m and the N test items were randomly partitioned into k non-overlapping samples of n items where $M = km$ and $N = kn$. The $k > 1$ item samples were randomly assigned and administered to the k examinee samples.

The terms "universe", "population" and "sample" are used in a hierarchical sense. A sample is a subset of a population which in turn is a subset of a universe.

The study differed from previously reported matrix sampling studies (Sjögren, 1970) along several dimensions: (1) the item population was a proper subset of a well-defined item universe, (2) the sampled items were administered not in a larger testing context but as an independent unit, (3) the examinee universe included relatively immature individuals in terms of mathematical and chronological maturity, and (4) the data generated was used to estimate group performance and to establish group norm statistics.

The specific questions considered in this report, a subset of a larger investigation (Montague, 1971) are the following:

1. Within an examinee population at each grade level and

    1.1 for a given item population and a given test format, do examinee population (Mean estimated test scores) differ when systematic distribution of test items to forms is contrasted with random assignment of items to forms?

    1.2 for a given item universe and a given test format, do examinee population METSs differ when systematic selection and systematic assignment of items to forms is contrasted with random selection and random assignment of items to forms?

    1.3 for a given item universe and a given test format, do examinee population METSs differ when systematic selection and random assignment of items to forms is contrasted with random selection and random assignment of items to forms?

    1.4 for a given item population, do examinee population METSs differ when systematic distribution of test items to k forms is contrasted with random assignment of items to 2k forms?

    1.5 for a given item universe, do examinee population METSs differ when systematic selection and systematic assignment of items to k forms is contrasted with random selection and random assignment of items to 2k forms?

2. Across examinee populations at each grade level and

    2.1 for a given item population and a given test format, do examinee population METSs differ when systematic distribution of test items to forms is contrasted with random assignment of items to forms?

    2.2 for a given item universe and a given test format, do examinee population METSs differ when systematic selection and systematic assignment of items to forms is contrasted with random selection and random assignment of items to forms?

    2.3 for a given item universe and a given test format, do examinee population METSs differ when systematic selection and random assignment of items to forms is contrasted with random selection and random assignment of items to forms?

2.4 for a given item population, do examinee population METSs differ when systematic distribution of test items to k forms is contrasted with random assignment of items to 2k forms?

2.5 for a given item universe, do examinee population METSs differ when systematic selection and systematic assignment of items to k forms is contrasted with random selection and random assignment of items to 2k forms?

Methods

Item Universe and Item Population: The item universe was the composite of two distinct domains: a domain of basic addition and subtraction facts and a domain of senctence types. Basic addition and subtraction facts whose sum was between 10 and 18 and whose addends were unequal comprised the domain of number facts. The domain of sentence types included simple equations in one unknown derived from $a \pm b = c$ and $c = a \pm b$. The 640-element item universe reflected a systematic assignment of each of the number facts to each of the simple equations.

Four 32-element proper subsets of the item universe were identified as item populations. Three of the item populations consisted of items randomly sampled from the item universe with inter-population item replacement and without intra-population item replacement. The fourth item population was identical to that used in a preexisting inventory (Weaver, 1970). The item populations are identified in Table 1.

Instruments: Seven test instruments reflecting one of four item populations were constructed. Test T7 consisted of items systematically drawn from the item universe and systematically assigned to four test forms each of eight items. The item population for T7 as well as the format was identical to that of the Weaver Inventory. The item populations for tests T1, T2, and

T3 consisted of all items from the item population for T7 randomly assigned to four or eight or two forms of 8 or 4 or 16 items respectively. The remaining tests, T4, T5, and T6, consisted of items randomly sampled from the item universe and randomly assigned to four, or eight, or two forms of 8, or 4, or 16 items respectively. The structure for the seven tests is contained in Table 2. A sample test precedes the references for this report.

Examinee Universe and Examinee Population: The examinee universe consisted of pupils using a common basal mathematics series from public schools in Madison, Wisconsin. From a listing of the 218 second and third grade classes, 79 classroom units were randomly selected. Across these units, 12 examinee populations at each grade were identified.

At each grade level, six examinee populations took form 1, or form 2 or form 3, or form 4 of the preexisting inventory and form 1, or form 2, ... , or form k ($k = 2, 4, 8$) of an item sampled test. Also at each grade level, six examinee populations took form 1, or form 2, ... , form k of an item sampled test. The characteristics of the examinee populations are identified in Table 3.

Data Source: The data unit within each examinee population was the estimated test score for a given examinee within that population derived from performance on one of the forms of a given test. For a given test of k forms each of n items per form such that $nk = 32$, the estimated test score for a given examinee was kn' where n' was the number of items correct. The mean estimated test score (METS) for a particular examinee population was the average of the estimated test scores for that population. In subsequent analyses, the data unit was the population METS: i.e., the METS based on all pupils within a given population who (regardless of classroom unit) took a particular test (regardless of the form of that test). One way ANOVAs were used with the entry being the population METS.

Table 4 contains the METS and the estimated test score variability within examinee populations for each test administered.

Results: No F-statistics were significant at the $\alpha = .05$ level of significance for questions 1.1, ... 1.5, 2.1 ... 2.5 as considered in this report. The results are summarized in Table 5 in terms of specific contrasts.

Educational Importance: Subject to the constraints of the present investigation the results suggest that adequate approximations of population METS were attainable for a variety of school examinee populations when the item universe and the examinee universe were well-defined and randomly sampled. Aspects of the feasibility include:

1. The procedure is appropriate when administering a sample of items in a separate testing context rather than as part of a larger testing program.

2. Wide diversity in test construction can be accommodated. The necessity for a single "standardized" instrument has been considerably reduced because comparable METSs can be obtained either by item sampling from an existing item population or from an encompassing item universe and randomly assigning items to the same number of forms or to fewer forms.

3. The efficiency of matrix sampling in terms of the time unit per pupil per item was supported. Accurate estimates of group achievement for relatively young children can be obtained in half the time per pupil. The procedure was empirically efficient for test forms with as few as four items.

4. To support or disclaim a contention that two groups were comparable with respect to selected, well-defined areas of content, alternative procedures have been provided. The alternatives provided for generalization to a well-defined universe of items, or to the population of items, or to the well-defined universe of examinees.

TABLE 1

Item Populations Used for Tests T1, T2, T3, T4, T5, T6, & T7

Item Population Tests T1, T2, T3, & T7*

$11 - x = 5$
$4 + 8 = x$
$9 = x - 8$
$5 - 12 = x$
$16 = 9 + x$
$x + 4 = 11$
$9 = x + 14$
$x = 13 - 9$

$x = 6 + 7$
$12 - 9 = x$
$x + 13 = 5$
$3 + x = 11$
$9 = 11 - x$
$x - 7 = 8$
$15 = x + 6$
$x = 8 - 14$

$13 = x + 4$
$8 + 9 = x$
$x = 11 - 7$
$9 - x = 14$
$6 + x = 11$
$5 = 12 + x$
$x - 9 = 7$
$8 = 12 - x$

$7 = x - 6$
$11 = 2 + x$
$14 + x = 8$
$15 - x = 6$
$x + 3 = 12$
$13 = 5 - x$
$x = 7 + 8$
$11 - 3 = x$

Item Population Test T4**

$x = 6 + 7$
$8 + 3 = x$
$5 = 12 - x$
$x - 6 = 5$
$13 = x + 4$
$7 - x = 13$
$x = 11 - 8$
$12 + x = 3$

$13 = 7 + x$
$5 = x - 9$
$3 = x - 9$
$8 + x = 15$
$2 + x = 11$
$x = 9 + 3$
$2 + 9 = x$
$14 + x = 5$

$x = 13 - 5$
$x + 13 = 9$
$x = 5 + 9$
$x + 11 = 2$
$x - 6 = 8$
$7 = 16 - x$
$x = 11 - 6$
$15 - 9 = x$

$12 - x = 7$
$11 = 9 - x$
$8 = 14 + x$
$14 = x + 8$
$9 - 15 = x$
$x = 3 + 8$
$x = 7 + 8$
$14 = x + 9$

Item Population Test T5**

$x - 7 = 8$
$13 - 9 = x$
$7 - x = 11$
$4 = x - 8$
$7 + 6 = x$
$12 - 5 = x$
$7 = x - 8$
$13 = 9 + x$

$3 = x - 9$
$3 - x = 12$
$11 - x = 9$
$4 + 8 = x$
$16 - x = 9$
$x = 16 - 9$
$9 = x + 17$
$x + 4 = 13$

$9 = x + 14$
$15 - x = 6$
$x - 8 = 5$
$15 = x + 8$
$x = 11 - 4$
$9 - x = 14$
$8 = x - 7$
$x = 9 + 3$

$12 = 4 + x$
$14 - x = 9$
$7 = 16 - x$
$11 - 4 = x$
$9 + x = 15$
$9 = x + 13$
$5 = 11 - x$
$8 = 13 + x$

---

*Items identical to those in preexisting inventory.
**Items sampled from item universe.

TABLE 1 continued

Item Populations Used for Tests
T1, T2, T3, T4, T5, T6, & T7

Item Population Test T6**

$x = 7 + 4$
$x + 7 = 11$
$6 - x = 14$
$15 = 7 - x$
$3 = x + 11$
$x + 7 = 13$
$x = 11 - 5$
$11 - 9 = x$
$8 = 14 - x$
$x - 7 = 8$
$9 + 3 = x$
$9 = 11 - x$
$12 = x + 8$
$x - 3 = 8$
$8 + x = 12$
$13 - x = 6$
$x + 4 = 11$
$4 - x = 12$
$x + 8 = 17$
$x = 3 + 9$
$7 - x = 15$
$6 - 13 = x$
$8 = 14 + x$
$11 = 7 + x$
$16 = 9 + x$
$x - 6 = 5$
$x + 4 = 13$
$11 = x + 8$
$5 = x - 8$
$x + 9 = 13$
$x - 5 = 9$
$15 + x = 7$

TABLE 2

Summary of Distinguishing Characteristics
Of Tests T1, T2, T3, T4, T5, T6, & T7

Test	Number of Items	Number of Forms	Number Items Per Form	Item Population Source	Item Population Selection Procedure	Item Assignment Procedure
T1	32	4	8	Weaver Inventory	Systematic	Random
T2	32	8	4	Weaver Inventory	Systematic	Random
T3	32	2	16	Weaver Inventory	Systematic	Random
T4	32	4	8	Item Universe	Random	Random
T5	32	8	4	Item Universe	Random	Random
T6	32	2	16	Item Universe	Random	Random
T7 [Weaver Inventory]	32	4	8	Item Universe	Systematic	Systematic

TABLE 3

Identification and Distribution of Examinee Populations E and F For Grades 2 and 3

Ss administered T7 <u>and</u> T1, T2, T3, T4, T5, or T6.	E1: 60/62*	E2: 76/70	E3: 63/61	E4: 59/64	E5: 67/71	E6: 47/64
Ss administered T1, T2, T3, T4, T5, or T6.	F1: 56/63	F2: 63/63	F3: 53/52	F4: 57/63	F5: 58/59	F6: 55/53
	T1: 4 forms 8 items each	T2: 8 forms 4 items each	T3: 2 forms 16 items each	T4: 4 forms 3 items each	T5: 8 forms 4 items each	T6: 2 forms 16 items each
	Items Sampled From Preexisting Inventory*			Items Sampled From Item Universe		

---

*The first numeral in each cell of the matrix indicates the number of second grade pupils in a particular population. The second numeral indicates the number of third grade pupils.

TABLE 4

Mean Estimated Test Scores and Estimated Test Score Variability Across Examinee Populations

Test	Number of Test Forms	Grade 2		Grade 3	
		Mean Estimated Test Score	Estimated Test Score Variability	Mean Estimated Test Score	Estimated Test Score Variability
Examinee Population E					
T1	4	19.79	8.08	23.61	5.96
T2	8	22.10	9.16	23.08	7.94
T3	2	20.82	6.11	23.60	3.17
T4	4	21.29	6.03	23.18	5.25
T5	8	18.94	9.29	22.42	7.31
T6	2	18.76	8.40	24.03	5.80
T7	4	19.99	6.52	22.99	5.94
Examinee Population F					
T1	4	19.21	8.26	22.85	7.65
T2	8	20.69	8.86	23.61	8.48
T3	2	20.45	7.54	22.96	6.74
T4	4	19.64	7.83	21.58	7.69
T5	8	17.92	8.18	21.83	9.06
T6	2	21.16	7.12	21.54	5.80

Note.--The mean estimated test score for a particular examinee population was

$$\frac{32\sum^{n} y_j}{n}$$

where y_j is the proportion correct for examinee j on one of the k forms of test Ti ($1 \leq i \leq 7$),

and n is the number of examinees within a particular population taking a form of Ti.

The estimated test score variability is the standard deviation of the distribution of the estimated test scores for the particular sample defined by the parameters of a particular contrast.

TABLE 5

Summary of Contrasts and Significance Levels

Question*	Contrast	Significance Level of Computed F	
		Grade 2	Grade 3
	Within Examinee Population E		
1.1	$\mu(T7) - \mu(T1) = 0$	$p > .05$	$p > .05$
1.2	$\mu(T7) - \mu(T4) = 0$	$p > .05$	$p > .05$
1.3	$\mu(T1) - \mu(T4) = 0$	$p > .05$	$p > .05$
1.4	$\mu(T7) - \mu(T2) = 0$	$p > .05$	$p > .05$
1.5	$\mu(T7) - \mu(T5) = 0$	$p > .05$	$p > .05$
	Across Examinee Populations E and F		
1.1	$\mu(T7,E) - \mu(T1,F) = 0$	$p > .05$	$p > .05$
1.2	$\mu(T7,E) - \mu(T4,F) = 0$	$p > .05$	$p > .05$
1.3	$\mu(T1,F) - \mu(T4,F) = 0$	$p > .05$	$p > .05$
1.4	$\mu(T7,E) - \mu(T2,F) = 0$	$p > .05$	$p > .05$
1.5	$\mu(T7,E) - \mu(T5,F) = 0$	$p > .05$	$p > .05$

*See pages 2 and 3 of this report.

My name is ______________________________

Grade ______ School ______________________________

0 1 2 ■ 4 5 6 7 8 9

10 11 12 13 14 15 ■ 17 18 19

■ . . . . . .

NUMBER PUZZLES

a. 7 - 2 = ■ ______

b. 9 + ■ = 10 ______

c. ■ = 3 + 5 ______

d. 0 - 4 = ■ ______

STOP

a. 13 = 7 + ■ ______

b. 5 = ■ - 9 ______

c. 3 = ■ - 9 ______

d. 8 + ■ = 15 ______

e. 2 + ■ = 11 ______

f. ■ = 9 + 3 ______

g. 2 + 9 = ■ ______

h. 14 + ■ = 5 ______

03 _ _ _ _ _ _ 04 02 _ _ _ _

References

Montague, M. A. Use of Matrix Sampling Procedures with Selected Examinee and Item Populations to Assess Achievement in Mathematics. Unplublished Ph.D. dissertation, University of Wisconsin, 1971.

Sjogren, D. D. Measurement Techniques in Evaluation. Review of Educational Research, 1970, 40, 301-320.

Weaver, V. F. Some Factors Associated with Pupils' Performance Levels on Simple Open Addition and Subtraction Sentences. The Arithmetic Teacher, 1971, 18, 513-519.