DOCUMENT RESUME

ED 065 506	24	TM 000 919

AUTHOR	Shoemaker, David M.
TITLE	Item-Examinee Sampling Procedures and Associated Standard Errors in Estimating Test Parameters.
INSTITUTION	Southwest Regional Educational Lab., Inglewood, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	TR-27
BUREAU NO	BR-6-2865
PUB DATE	5 Jun 70
NOTE	21p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Error Patterns; *Item Sampling; Research; *Sampling; Statistical Analysis; *Test Interpretation; *Test Results

ABSTRACT

Selected parameters for a negatively-skewed and a normally distributed normative distribution were estimated in a post-mortem item-examinee sampling investigation. Manipulated systematically were number of subtests, number of items per subtest, and number of examinees responding to each subtest. Each item-examinee sampling procedure was replicated five times. Defining one observation as the score received by one examinee on one item, the results indicate that the mean of a normative distribution is easily and efficiently estimated with a relatively small number of observations; the variance, to the contrary is a more difficult parameter to approximate and requires a larger number of observations to obtain a reasonable efficient estimate. The results of this investigation support the conclusion that, in estimating parameters by item-examinee sampling, the variable of importance is not the item-examinee sampling procedure but is instead the number of observations obtained by that procedure. (Author)

OE-NCERD
TM
BR-6-2865
PA-24

SWRL

SOUTHWEST REGIONAL LABORATORY FOR EDUCATIONAL RESEARCH & DEVELOPMENT

ED 065506

TM 000 919

Item-Examinee Sampling Procedures and Associated Standard Errors in Estimating Test Parameters

27 5 JUNE 1970

1

Published by Southwest Regional Laboratory for Educational Research and Development, a public agency supported as a regional educational laboratory by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

ITEM-EXAMINEE SAMPLING PROCEDURES AND ASSOCIATED STANDARD ERRORS IN ESTIMATING TEST PARAMETERS

David M. Shoemaker

ABSTRACT

Selected parameters for a negatively-skewed and a normally distributed normative distribution were estimated in a post-mortem item-examinee sampling investigation. Manipulated systematically were number of subtests, number of items per subtest, and number of examinees responding to each subtest. Each item-examinee sampling procedure was replicated five times. Defining one observation as the score received by one examinee on one item, the results indicate that the mean of a normative distribution is easily and efficiently estimated with a relatively small number of observations; the variance, to the contrary, is a more difficult parameter to approximate and requires a larger number of observations to obtain a reasonable efficient estimate. The results of this investigation support the conclusion that, in estimating parameters by item-examinee sampling, the variable of importance is not the item-examinee sampling procedure but is instead the number of observations obtained by that procedure.

ITEM-EXAMINEE SAMPLING PROCEDURES AND ASSOCIATED STANDARD ERRORS IN ESTIMATING TEST PARAMETERS

An important aspect of large-scale tryouts of criterion-referenced instructional programs is the collection of student achievement data indicating the effectiveness of the program. Collection of this data frequently involves individual administration of criterion-referenced tests--a procedure which is time-consuming and costly to implement with the entire tryout population in a large-scale tryout. However, accurate estimates of the population mean and variance can be obtained through item-examinee sampling, a much more economical procedure. The study described herein was conducted to investigate the utility of various item-examinee sampling procedures when used for group assessment with criterion-referenced instructional programs.

Item-examinee Sampling

Item-examinee sampling is a procedure in which a set of K test items is subdivided into t subtests of items and each subtest of items is administered to different subgroups of examinees selected from the testable population of N examinees. Although each examinee receives only a proportion of the complete set of items, the statistical model described by Lord (1960, 1962) permits the researcher to estimate the mean and variance of the total test score distribution which would have been obtained by testing N examinees on K items. To demonstrate the procedure and applicability of item-examinee sampling in educational research, consider the following situation: A 100-item comprehensive examination is to be administered to 5000 grade 1 students at the end of a specific instructional program. The purpose of the examination is that of group assessment, not individual assessment. For various reasons, e.g., it is not economically feasible to administer the complete set of items to all examinees, the amount of testing time is prohibitive, the scoring costs are prohibitive, or the cooperation of individual schools could be more readily obtained if only a few minutes of each student's time were required, item-examinee sampling is a desirable experimental procedure. One possible item-examinee sampling procedure which might be used in this situation is as follows: (a) The 100-item test is subdivided into five subtests each containing 20 items. Items are assigned to subtests by sampling at random and without replacement from the 100-item pool. (b) Each subtest is administered to three classes of examinees which have been, for each subtest, randomly selected without replacement from the pool of testable classes. In this particular procedure, approximately 450 students would be tested (assuming 30 students per class) over 20 items; however, not all students would receive the same 20 items. The testing time per examinee would be approximately 1/5 of the time required to administer 100 items.

The mean and variance for the total test are estimated from subtest results by

$$\hat{\mu}_i = \frac{K}{k_i}\,\bar{X}_i \text{ and}$$

$$\hat{\sigma}_i^2 = \frac{m_i\,K[(K-1)\,s_i^2 - (K-k_i)\sum^{k_i} p_i q_i]}{k_i(k_i-1)(m_i-1)},$$

where, referring to the i th subtest,

K is the number of items in the complete test,

m_i is the number of examinees taking the subtest,

k_i is the number of subtest items,

$\bar{X}_i$ is the mean subtest score,

s_i^2 is the variance of the subtest scores, and

$\sum^{k_i} p_i q_i$ is the sum of the k_i subtest item variances.

A single estimate of μ is obtained by averaging the t estimates of μ obtained from each item-examinee sample; a single estimate of σ^2, by averaging the t estimates of the population variance. If the total number of examinees N is less than 500, the pooled estimate of σ^2 is multiplied by $(N-1)/N$.

Item-examinee sampling differs from item-sampling and from examinee-sampling. In item-sampling, a randomly selected subset of test items is administered to all examinees; in examinee-sampling, all items are administered to a randomly selected subgroup of examinees. Both item-sampling and item-examinee sampling procedures implicitly assume that examinee performance on an item does not depend on the context in which the item occurs. This is a critical assumption and must be evaluated carefully in each situation. It must be emphasized that item-examinee sampling is a group assessment procedure.

Procedural Guidelines in Item-examinee Sampling

While it is undeniably true that item-examinee sampling is an effective norming technique, few procedural guidelines are available to aid the researcher in determining the most appropriate number of subtests, number of items per subtest, and number of examinees per subtest to use in an item-examinee sampling investigation. Shoemaker (1970), using a post-mortem item-examinee sampling paradigm, manipulated systematically the variables of number of subtests, number of items per subtest, and number of examinees responding to each subtest in determining the most appropriate procedure to use when estimating a normative distribution. Defining one observation as the score received by one examinee on one item, the results suggested that as the number of observations increased beyond 1,500 all item-examinee sampling procedures produce distributions stochastically equivalent to the normative distribution. Shoemaker concluded that, in estimating a norm distribution by item-examinee sampling, the variable of importance is not the item-examinee sampling procedure per se but is instead the number of observations obtained by that procedure.

The investigation described herein was designed to isolate those factors which produced the Shoemaker (1970) results. Major considerations were as follows: (1) The distribution of test scores in the Shoemaker investigation was normal and it is possible that item-examinee sampling as a technique may be robust for normal distributions. Distribution parameters should be estimated by a multitude of item-examinee sampling procedures when the normative distribution is not normal. (2) Results of 15 item-examinee sampling procedures were reported by Shoemaker. Each procedure produced a pooled estimate of the population mean μ and a pooled estimate of the population variance σ^2. While the procedure used in item-examinee sampling was not found to be a significant factor, one sampling procedure may be preferred to another if estimates of test parameters resulting from that procedure have less variance than corresponding estimates obtained from another procedure. Thus, standard errors of estimate per item-examinee sampling procedure (not computed in the Shoemaker investigation) should be determined empirically for a wide variety of item-examinee sampling procedures. (3) In the majority of item-examinee sampling investigations, sampling of items has been exhaustive and without replacement, that is, all test items have appeared in the subtests and no item was included in more than one subtest. Lord and Novick (1968, p. 257) have indicated that failure to administer all test items inflates the standard error of estimating the population mean by item-examinee sampling. Furthermore, the smaller the number of items in the population, the worse the effect. A statement such as this is easily verified empirically and its generalizability to estimating the population variance should also be considered. It may be hypothesized that improved estimates of parameters are obtained if a particular item appears in more than one subtest. Considerations such as these served as the basis for the experimental manipulation described herein. The specific parameters to be estimated were the mean μ, variance σ^2, and Kuder-Richardson Formula 21 reliability coefficient for the normative distribution of total test scores.

Method

The research design was one of post-mortem item-examinee sampling: given a normative distribution, various item-examinee samples are randomly selected from this data base and used to estimate parameters of the distribution from which they have been sampled. The first (of two) normative distributions considered consisted of test scores received by 1,031 kindergarten students on a 20-item dichotomously-scored three-alternative multiple-choice criterion-referenced examination administered during the fall of 1969 as part of the First-Year Communication Skills Program at the Southwest Regional Laboratory for Educational Research and Development (SWRL). The descriptive statistics for this markedly negatively-skewed distribution are given in column 3 in Table 1.

The 36 item-examinee sampling procedures used are described in the first five columns of Table 2. Three levels of number of subtests (10, 5, 2), three levels of number of items per subtest (15, 10, 5), and four levels of number of examinees per subtest (120, 90, 60, 30) were manipulated systematically. The results obtained from each of the 36 item-examinee sampling procedures were replicated five times.

Results I

The results of the 36 item-examinee sampling procedures are given in columns 6 through 9 in Table 2. As all procedures are similar, only the procedure and results outlined in the first row of Table 2 will be described in detail. In the first item-examinee sampling procedure, 10 subtests, each containing 15 items were formed. To have 15 items per subtest, items had to be sampled for each subtest with replacement (WR) from the 20 item population. Each subtest was administered to 30 examinees sampled without replacement (WOR) from the testable population of 1,031 examinees. Each item-examinee sampling procedure produced one pooled estimate of μ and one pooled estimate of σ^2. As each sampling procedure had been replicated five times, there were five estimate of μ and five estimates of σ^2. In the first item-examinee sampling procedure, the mean of the five estimates of μ was 17.571; the standard deviation of these five estimates (or the standard error of estimate in estimating the population mean associated with the first procedure) was .105. The mean of the five

TABLE 1

DESCRIPTIVE STATISTICS FOR NORMAL AND SKEWED NORMATIVE DISTRIBUTIONS

Test Score	Frequency	
	Normal Dist.	Skewed Dist.
0	12	0
1	14	0
2	24	1
3	36	0
4	45	0
5	42	3
6	71	5
7	55	5
8	94	2
9	90	7
10	88	19
11	91	15
12	74	29
13	84	31
14	56	40
15	55	46
16	35	48
17	22	85
18	18	168
19	16	207
20	9	320
Number of Examinees	1,031	1,031
Mean Test Score	9.840	17.543
Variance of Test Scores	18.889	8.950
KR21	.774	.799

TABLE 2

ITEM-EXAMINEE SAMPLING PROCEDURES WITH RESULTS: SKEWED DISTRIBUTION

Procedure	Number of Subtests	Number of Items per Subtest	Number of Examinees per Subtest	Item Sampling Plan	Examinee Sampling Plan	$\bar{\hat{\mu}}$	$\bar{\hat{\sigma}}^2$	SE($\hat{\mu}$)	SE($\hat{\sigma}^2$)	Total Number of Obs.	Average No. of Items Omitted
1	10	15	30	WR	WOR	17.571	9.612	.105	.497	4500	0
2	10	10	30	WR	WOR	17.539	8.12[illegible]	.393	.442	3000	0
3	10	5	30	WR	WOR	17.455	10.734	.131	1.009	1500	0.8
4	10	15	60	WR	WOR	17.555	9.029	.090	.646	9000	0
5	10	10	60	WR	WOR	17.618	8.475	.161	.892	6000	0
6	10	5	60	WR	WOR	17.721	9.111	.134	.431	3000	1.2
7	10	15	90	WR	WOR	17.558	8.969	.055	.456	13500	0
8	10	10	90	WR	WOR	17.504	8.982	.063	.321	9000	0
9	10	5	90	WR	WOR	17.548	9.431	.187	.493	4500	1.0
10	10	15	120	WR	WR	17.351	9.094	.032	.456	18000	0
11	10	10	120	WR	WR	17.448	8.922	.100	.744	12000	0
12	10	5	120	WR	WR	17.226	9.098	.167	1.201	6000	0.8
13	5	15	30	WR	WOR	17.385	9.021	.318	1.602	2250	0
14	5	10	30	WR	WOR	17.235	9.382	.235	1.744	1500	1.0
15	5	5	30	WR	WOR	17.371	9.590	.517	3.594	750	3.4
16	5	15	60	WR	WOR	17.278	9.336	.130	1.031	4500	0
17	5	10	60	WR	WOR	17.454	8.892	.170	.675	3000	0.6
18	5	5	60	WR	WOR	17.091	9.348	.263	1.692	1500	5.4
19	5	15	90	WR	WOR	17.420	8.871	.045	.574	6750	0
20	5	10	90	WR	WOR	17.399	9.165	.148	.415	4500	0.6

TABLE 2 (Continued)

ITEM-EXAMINEE SAMPLING PROCEDURES WITH RESULTS: SKEWED DISTRIBUTION

Procedure	Number of Subtests	Number of Items per Subtest	Number of Examinees per Subtest	Item Sampling Plan	Examinee Sampling Plan	$\bar{\hat{\mu}}$	$\bar{\hat{\sigma}}^2$	$SE(\hat{\mu})$	$SE(\hat{\sigma}^2)$	Total number of Obs.	Average No. of Items Omitted
21	5	5	90	WR	WOR	17.207	9.632	.281	.857	2250	5.0
22	5	15	120	WR	WOR	17.380	9.232	.118	.535	9000	0.2
23	5	10	120	WR	WOR	17.295	9.875	.138	.710	6000	0.8
24	5	5	120	WR	WOR	17.440	8.186	.195	.608	3000	4.8
25	2	15	30	WR	WOR	17.431	8.166	.550	1.805	900	2.4
26	2	10	30	WOR	WOR	17.200	10.686	.580	4.127	600	0
27	2	5	30	WOR	WOR	17.320	11.320	.565	3.394	300	10.0
28	2	15	60	WR	WOR	17.311	8.693	.192	1.280	1800	1.4
29	2	10	60	WOR	WOR	17.517	7.478	.167	.974	1200	0
30	2	5	60	WOR	WOR	17.053	8.573	.288	2.073	600	9.6
31	2	15	90	WR	WOR	17.413	9.074	.285	1.615	2700	1.0
32	2	10	90	WOR	WOR	17.335	8.728	.176	1.972	1800	0
33	2	5	90	WOR	WOR	17.187	7.845	.257	1.746	900	10.0
34	2	15	120	WR	WOR	17.510	8.425	.155	1.375	3600	0.8
35	2	10	120	WOR	WOR	17.482	9.135	.155	1.520	2400	0
36	2	5	120	WOR	WOR	17.510	8.057	.431	1.334	1200	10.0
Norm	1	20	1,031	WOR	WOR	17.543	8.950			20620	

estimates of σ^2 was 9.612; their standard deviation was .497. Each replication involved 4500 = (10)(15)(30) observations. As indicated in the last column in Table 2, all test items were included in one or another of the 10 subtests.

The results in Table 2 could be (and have been) rearranged in a number of ways; however, the most meaningful way appeared to be by number of observations. This has been done and is found in columns 3 through 6 in Table 3. A graphic display of the same results is given in Figure 1.

The estimate of $\hat{\mu}$ and $\hat{\sigma}^2$ obtained in each replication was used to compute $\widehat{KR21}$ as an estimate of the KR21 obtained using parameters. The mean coefficient across replications and standard error per item-examinee sampling procedure are given in columns 2 and 3 in Table 4.

Degree of Skewness in Normative Distribution

It is not unreasonable to hypothesize that these results may be due to the extreme degree of skewness in the normative distribution. Would results differ if a normal normative distribution of test scores on a 20-item test had been used in place of the skewed distribution? To answer questions such as this and, more directly, to investigate the effect of degree of skewness in the normative distribution on standard errors of item-examinee sampling procedures, all item-examinee sampling procedures were replicated using a normal normative distribution.

Item scores for 1,031 examinees on a 20-item test were generated by a Monte Carlo approach such that the distribution of total test scores was normal with item difficulty indices (proportion of examinees answering each item correctly) approximately equal to .5 and the Kuder-Richardson Formula 21 reliability of the total test being .774. Descriptive statistics for the normal normative distribution are given in column 2 in Table 1.

All item-examinee sampling procedures were repeated with the normal distribution serving as the normative distribution. The statistical analyses were identical to those reported for the skewed distribution case.

Results II

The results of the 36 item-examinee sampling procedures are given in columns 6 through 9 in Table 5. The results in Table 5 are

TABLE 3

MEAN RESULTS PER ITEM-EXAMINEE SAMPLING PROCEDURE ORGANIZED ACCORDING TO NUMBER OF OBSERVATIONS FOR NORMAL AND SKEWED NORMATIVE DISTRIBUTIONS

Number of Observations		Skewed Normative Distribution				Normal Normative Distribution			
		$\bar{\hat{\mu}}$	SE($\hat{\mu}$)	$\bar{\hat{\sigma}}^2$	SE($\hat{\sigma}^2$)	$\bar{\hat{\mu}}$	SE($\hat{\mu}$)	$\bar{\hat{\sigma}}^2$	SE($\hat{\sigma}^2$)
18000	(1)[a]	17.351	.032	9.094	.456	9.773	.089	19.212	1.110
13500	(1)	17.558	.055	8.969	.456	9.886	.055	18.720	.381
12000	(1)	17.448	.100	8.922	.744	9.825	.134	19.196	.750
9000	(3)	17.480	.095	9.081	.519	9.787	.077	18.907	.509
6750	(1)	17.420	.045	8.871	.574	9.812	.155	18.870	.738
6000	(3)	17.380	.155	9.149	.956	9.782	.084	18.662	1.046
4500	(4)	17.449	.145	9.386	.657	9.951	.249	18.755	.816
3600	(1)	17.510	.155	8.425	1.375	8.853	.141	18.046	1.014
3000	(4)	17.539	.245	8.479	.550	9.729	.239	18.527	1.447
2700	(1)	17.413	.285	9.074	1.615	9.907	.214	18.063	1.012
2400	(1)	17.482	.155	9.135	1.520	9.647	.105	17.883	1.781
2250	(2)	17.296	.300	9.327	1.284	9.667	.358	19.280	1.783
1800	(2)	17.323	.184	8.711	1.662	9.831	.307	18.954	1.657
1500	(3)	17.260	.217	9.821	1.519	9.746	.214	18.668	1.512
1200	(2)	17.514	.327	7.768	1.168	9.699	.400	18.213	2.799
900	(2)	17.309	.429	8.006	1.776	9.800	.424	19.218	3.511
750	(1)	17.371	.517	9.590	3.594	10.021	.361	18.808	1.526
600	(2)	17.127	.458	9.630	3.266	9.973	.167	16.727	3.291
300	(1)	17.320	.565	11.320	3.394	9.840	.856	15.605	3.935
Norm		17.543		8.950		9.840		18.889	

[a] Number of item-examinee sampling procedures pooled.

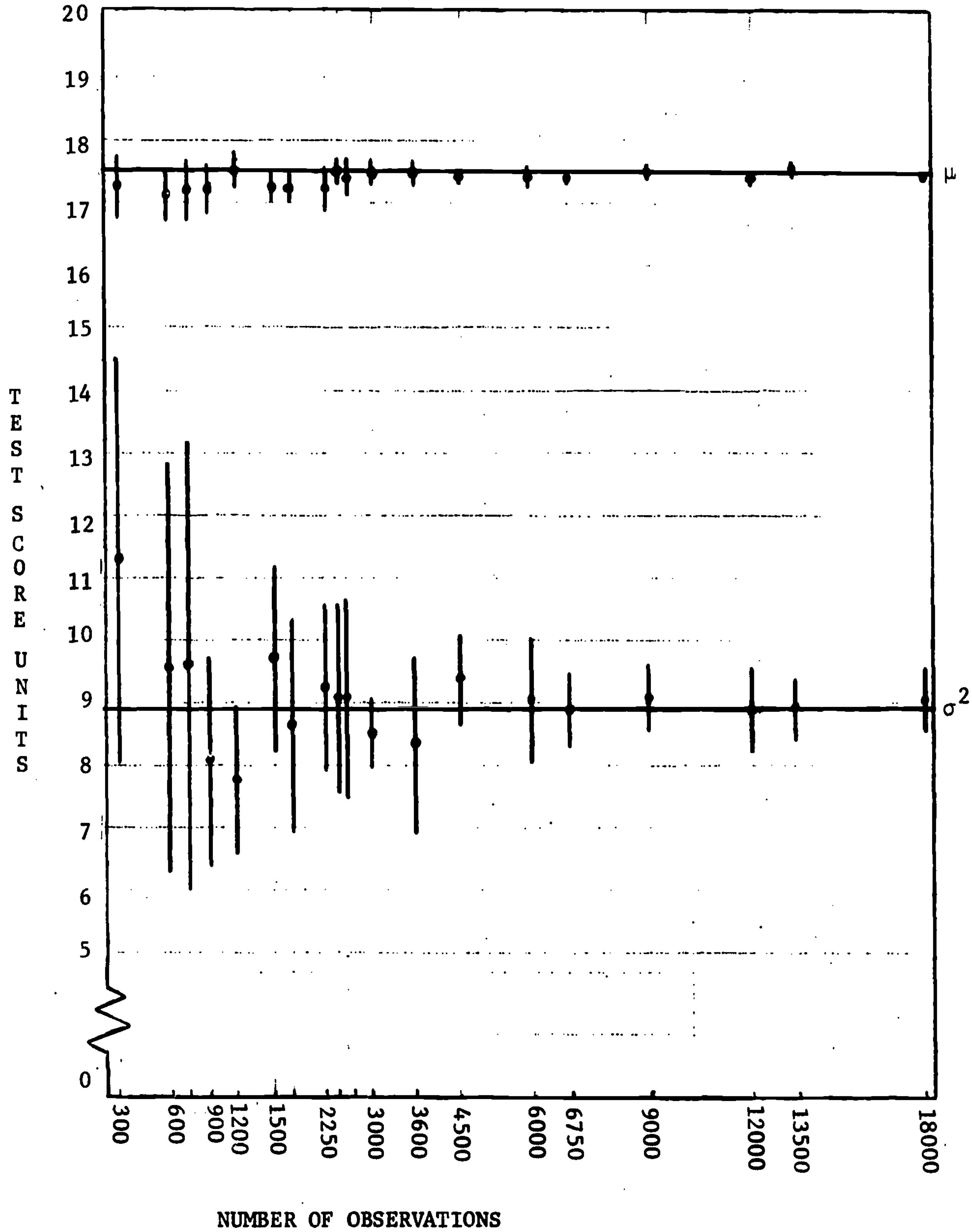

Figure 1: Mean estimate and ± one standard error of estimate for mean μ and variance σ^2 as a function of the number of observations for the skewed distribution case. Standard errors of estimate are based on five replications.

TABLE 4

MEAN KR21 COEFFICIENTS AND ASSOCIATED STANDARD ERRORS OF ESTIMATE
PER ITEM-EXAMINEE SAMPLING PROCEDURE
FOR SKEWED AND NORMAL NORMATIVE DISTRIBUTIONS

Item-Examinee Sampling Procedure	Skewed Distribution		Normal Distribution	
	$\overline{\widehat{KR21}}$	$SE(\widehat{KR21})$	$\overline{\widehat{KR21}}$	$SE(\widehat{KR21})$
1	.807	.006	.774	.019
2	.772	.015	.779	.013
3	.833	.020	.771	.034
4	.802	.010	.779	.006
5	.791	.016	.763	.021
6	.819	.003	.745	.029
7	.801	.006	.771	.006
8	.797	.006	.775	.063
9	.813	.017	.773	.011
10	.786	.010	.778	.017
11	.789	.020	.778	.010
12	.773	.028	.767	.006
13	.783	.032	.778	.028
14	.778	.046	.783	.013
15	.787	.045	.771	.025
16	.790	.027	.775	.006
17	.789	.014	.771	.028
18	.767	.039	.752	.018
19	.785	.020	.773	.011
20	.796	.008	.765	.013
21	.789	.026	.778	.025
22	.793	.008	.768	.011
23	.802	.013	.779	.017
24	.765	.016	.774	.017
25	.762	.034	.767	.023

TABLE 4 (Continued)

MEAN KR21 COEFFICIENTS AND ASSOCIATED STANDARD ERRORS OF ESTIMATE
PER ITEM-EXAMINEE SAMPLING PROCEDURE
FOR SKEWED AND NORMAL NORMATIVE DISTRIBUTIONS

Item-Examinee Sampling Procedure	Skewed Distribution		Normal Distribution	
	$\overline{\widehat{KR21}}$	$SE(\widehat{KR21})$	$\overline{\widehat{KR21}}$	$SE(\widehat{KR21})$
26	.773	.115	.687	.116
27	.811	.094	.694	.095
28	.767	.029	.779	.025
29	.743	.034	.730	.051
30	.732	.058	.752	.051
31	.788	.028	.761	.015
32	.765	.044	.767	.024
33	.714	.068	.777	.054
34	.775	.031	.760	.016
35	.794	.035	.756	.029
36	.768	.020	.782	.027
Norm	.799		.774	

TABLE 5

ITEM-EXAMINEE SAMPLING PROCEDURES WITH RESULTS: NORMAL DISTRIBUTION

Procedure	Number of Subtests	Number of Items per Subtest	Number of Examinees per Subtest	Item Sampling Plan	Examinee Sampling Plan	$\bar{\hat{\mu}}$	$\bar{\hat{\sigma}}^2$	SE($\hat{\mu}$)	SE($\hat{\sigma}^2$)	Total Number of Obs.	Average No. of Items Omitted
1	10	15	30	WR	WOR	10.132	18.931	.326	1.194	4500	0
2	10	10	30	WR	WOR	9.964	19.104	.322	.898	3000	0
3	10	5	30	WR	WOR	9.974	18.938	.156	2.152	1500	0.8
4	10	15	60	WR	WOR	9.781	19.246	.074	.409	9000	0
5	10	10	60	WR	WOR	9.754	18.263	.106	1.282	6000	0
6	10	5	60	WR	WOR	9.620	17.273	.081	1.833	3000	1.2
7	10	15	90	WR	WOR	9.886	18.720	.054	.380	13500	0
8	10	10	90	WR	WOR	9.838	18.956	.097	.391	9000	0
9	10	5	90	WR	WOR	9.846	18.820	.114	.683	4500	1.0
10	10	15	120	WR	WR	9.773	19.212	.088	1.109	18000	0
11	10	10	120	WR	WR	9.825	19.196	.134	.751	12000	0
12	10	5	120	WR	WR	9.740	18.447	.033	.305	6000	0.8
13	5	15	30	WR	WOR	9.714	19.264	.464	1.756	2250	0
14	5	10	30	WR	WOR	9.675	19.509	.287	.943	1500	1.0
15	5	5	30	WR	WOR	10.021	18.808	.361	1.527	750	3.4
16	5	15	60	WR	WOR	9.998	18.945	.276	.348	4500	0
17	5	10	60	WR	WOR	9.633	18.819	.308	1.701	3000	0.6
18	5	5	60	WR	WOR	9.589	17.557	.180	1.156	1500	5.4
19	5	15	90	WR	WOR	9.812	18.870	.153	.737	6750	0
20	5	10	90	WR	WOR	9.827	18.324	.232	.808	4500	0.6

TABLE 5 (Continued)

ITEM-EXAMINEE SAMPLING PROCEDURES WITH RESULTS: NORMAL DISTRIBUTION

Procedure	Number of Subtests	Number of Items per Subtest	Number of Examinees per Subtest	Item Sampling Plan	Examinee Sampling Plan	$\bar{\hat{\mu}}$	$\bar{\hat{\sigma}}^2$	SE($\hat{\mu}$)	SE($\hat{\sigma}^2$)	Total Number of Obs.	Average No. of Items
21	5	5	90	WR	WOR	9.619	19.295	.202	1.808	2250	5.0
22	5	15	120	WR	WOR	9.742	18.519	.064	.675	9000	0.2
23	5	10	120	WR	WOR	9.852	19.276	.086	1.118	6000	0.8
24	5	5	120	WR	WOR	9.697	18.911	.147	1.147	3000	4.8
25	2	15	30	WR	WOR	10.035	18.467	.550	1.524	900	2.4
26	2	10	30	WOR	WOR	9.993	15.443	.216	3.426	600	0
27	2	5	30	WOR	WOR	9.840	15.605	.855	3.935	300	10.0
28	2	15	60	WR	WOR	9.906	19.344	.398	1.752	1800	1.4
29	2	10	60	WOR	WOR	9.670	16.808	.334	3.355	1200	0
30	2	5	60	WOR	WOR	9.953	18.010	.098	3.149	600	9.6
31	2	15	90	WR	WOR	9.907	18.063	.214	1.012	2700	1.0
32	2	10	90	WOR	WOR	9.756	18.564	.171	1.556	1800	0
33	2	5	90	WOR	WOR	9.565	19.968	.240	4.725	900	10.0
34	2	15	120	WR	WOR	9.853	18.046	.140	1.014	3600	0.8
35	2	10	120	WOR	WOR	9.647	17.883	.103	1.781	2400	0
36	2	5	120	WOR	WOR	9.727	19.618	.456	2.101	1200	10.0
Norm	1	20	1,031	WOR	WOR	9.840	18.889			20620	

interpreted in the same manner as those in Table 2. Results for KR21 coefficients are given in columns 4 and 5 in Table 4.

The results in Table 5 have been rearranged and pooled according to number of observations and are given in columns 7 through 10 in Table 3. A graphic display of the same results is given in Figure 2.

Discussion

The most apparent difference in standard errors associated with item-examinee sampling procedures is the difference in magnitude between the standard error in estimating the population mean and the standard error in estimating the population variance. Item-examinee sampling procedures are generally efficient in estimating μ and much less efficient in estimating σ^2 . Indeed, it would seem that almost any procedure could be used in estimating μ. In view of these results, it is not surprising that all item-examinee sampling investigations in the literature have reported satisfactory estimates of the mean of the normative distribution. The degree of accuracy in estimating σ^2 is most obviously a function of the number of observations. Parameters for both distributions can be estimated accurately given a large number of observations; the number of observations taken by the researcher should be determined by the choice of parameter to be estimated and the desired accuracy of the results. Results do not appear to be influenced significantly by degree of skewness in the normative distribution.

For several of the item-examinee sampling procedures considered, e.g., number 1 in Table 2, identical items were included in more than one subtest. Other sampling procedures sampled items exhaustively and without replacement: all items were sampled and no item was included in more than one subtest. An example of this case is procedure 26 in Table 2. In some procedures, a number of items were excluded from all subtests as, for example, procedure 27 in Table 2. The effect of these sampling variations on the standard errors of estimate are most appropriately interpreted in terms of number of observations: the greater the number of observations, the less the standard error of estimate. If the results in Tables 2 and 5 are individually rearranged (without averaging results over procedures having the same number of observations) and standard errors are examined as a function of number of items omitted, no trend is apparent. If failure to administer all test items does influence standard errors of estimate, perhaps the effect would have been more apparent if the number of replications per item-examinee sampling procedure had been significantly increased.

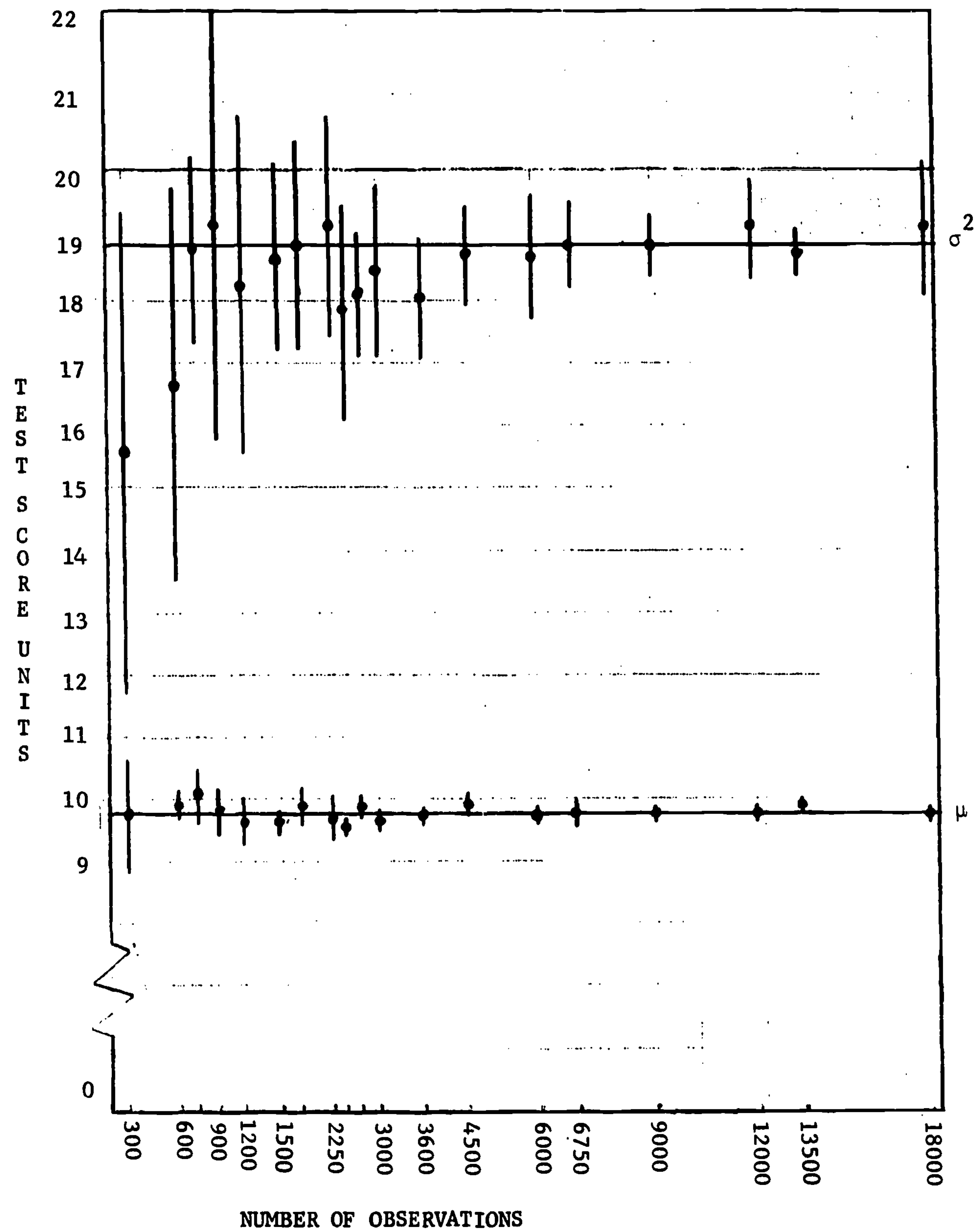

Figure 2: Mean estimate and $\pm$ one standard error of estimate for mean μ and variance σ^2 as a function of the number of observations for the normal distribution case. Standard errors of estimate are based on five replications.

In general, the results of this investigation support the conclusion that, in estimating parameters by item-examinee sampling, the variable of importance does not appear to be the item-examinee sampling procedure but is instead the number of observations obtained by that procedure. All item-examinee sampling, item-sampling, and examinee-sampling investigations should report standard errors for each sampling procedure considered. The interpretation of results is greatly simplified in the light of standard errors of estimate per parameter.

REFERENCES

Lord, F. M. Use of the true-score theory to predict moments of univariate and bivariate observed-score distributions. Psychometrika, 1960, 25, 325-342.

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

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

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