THE SAMPLE DESIGN FOR THE SURVEY OF THE INTERNATIONAL PROJECT ON THE EVALUATION OF EDUCATIONAL ACHIEVEMENT (I.E.A.) IN JAPAN

TATSUZO SUZUKI

(Received Aug. 5, 1964)

1. Introduction

Since 1960, the project of the cross-national study of educational achievement in mathematics has been carried out in 13 countries in the world under the auspices of the UNESCO Institute for Education, Hamburg, Germany [1].

Japan is participating in this project, and the National Institute for Educational Research has been conducting a national test and other related surveys.

We are cooperating with the project and made an efficient sample design.

This paper summarizes the research on the sample design used in the I.E.A. test-project in Japan. The design used here is that of a usual stratified two-stage random sampling, and the method of estimation is an unbiased one, because it is necessary for the whole project that samples drawn in different countries are comparable and are unbiased ones from the populations to be defined.

The purposes of our research are to provide advance estimates of the expected precision of results from samples of specified sizes and to estimate the gains due to the stratification by using related data in hand and to propose an efficient design incorporating these two characteristics.

The empirical results presented in this paper may be helpful to survey designs in related fields.

2. Populations and samples

1° The four samples

For the four surveys in this project different samples were needed, and these different samples were drawn from the following populations:

Level 1a. All pupils who are 13.0–13.11 years old at the time of testing.

Level 1b. All pupils at the grade level where a majority of the pupils are 13.0–13.11 years old.

The sizes of these samples are suggested to be 3,000∼5,000 pupils. If it is convenient, some pupils may be included in the age sample 1a and the grade sample 1b at the same time.

Level 2. All pupils at an intermediate grade level (or several grade levels) near the end of the comprehensive schooling.

The sample size is about 1,000.

Level 3. All pupils who are finishing a full-time course, and are normally preparing for university studies, or for other studies at this level, or for work at a similar level.

The selection of the point of testing and the definition of the population are matters to be decided by each country. The sample size is about 1,000.

2° Definitions of populations

In Japan, the level 1 corresponds to the second grade of the lower secondary (junior high) school, and the level 3 corresponds to the third grade of the upper secondary (senior high) school.

The lower secondary school is compulsory, and those who are not enrolled in lower secondary schools are at most 0.3% of this age-group. Pupils who are repeating the same grade due to illness, etc., are 1.2% in each of these grades.

Then, we used the following as the operational definitions of the populations.

Level 1. All pupils who are enrolled in the second grade of lower secondary school at the time of testing (April, 1964).

Level 3. All pupils who are enrolled in the third grade of upper secondary school at the time of testing.

The numbers of schools and of pupils in these grades were as follows as of May 1, 1962 :

Table 1. Number of schools and of pupils.

Level	Type of school		Number of schools (courses)	Number of pupils (1962)
Level 1	Public lower secondary school		12,030	2,298,166
	Private lower secondary school		617	80,641
Level 3	Public upper secondary school	General course	1,599	427,618
		Technical course	369	83,218
		Commercial course	573	104,513
		Agricultural course	480	58,667
		Home economics course	666	45,024
		Others	69	6,833
	Private upper secondary school	General course	836	242,080
		Commercial course	388	89,777
		Home economics course	233	28,714
		Agricultural, Technical and Others	167	42,912

3° General principle of sampling procedures

In order to draw comparable samples from different countries, probability sampling must be used, and probably the most convenient procedure is that of stratified two-stage sampling.

The sampling unit of the first stage is school (or course) and that of the second stage is pupil.

At the first stage, after stratifying the schools (or the courses, in the case of upper secondary schools) so as to make the variance within each stratum as small as possible, one or a few schools are drawn from each stratum in proportion to the number of pupils in that stratum. We used the numbers of pupils found by the 1963 school year survey.

At the second stage, after classifying pupils in that grade of the selected schools by such characteristics as scores of an intelligence test or the year-end examination in mathematics, individual pupils are drawn from each selected school by using the systematic sampling method. The number of pupils drawn from each stratum is proportionate to that stratum size*).

4° Stratification materials

We studied materials which would make this design as efficient as possible. Data concerning schools which were available were those on

*) This method (size proportionate allocation) is often used in practical surveys, for analytical advantages.

the mathematics part in the national tests*⁾, the school average score in mathematics in the 1959 sample test; the school average score in mathematics in the 1961 national test; the type of district to which a school belong (see Appendix page 77); the school average score in mathematics in the 1962 national test; the percentage of those who desire to go to upper secondary schools against the total number of pupils in the same grade in a school; the enrollment of the school.

5° Stratification of schools (the first-stage sampling units)

The separate procedure of stratification was used for each of the following five categories of schools:

a) public lower secondary schools, b) private lower secondary schools, c) the general course of public upper secondary schools, d) the four courses: agricultural, technical, commercial and home economics of public upper secondary schools**⁾, e) the general and the other courses of private upper secondary schools.

The stratification procedures of public lower secondary schools (category a) are based on the average scores of mathematics of the above mentioned schools in the 1961 and 1962 national tests; the aspect of district to which each school belongs; ratio of those pupils who intend to go to upper secondary schools to the total enrollment; enrollment of the school. The strata were set up on the basis of the α-score calculated by a method of quantification, [3] (see section 3 page 74).

Stratifications for category c) (the general course of public upper secondary schools) are based on the aspect of district; ratio of those pupils who intend to go to college to the total enrollment of the school.

The stratifications for categories b), d) and e) are based on type of school, aspect of district, enrollment and the rating of schools by the expert***⁾.

The strata were then set up on the basis of subjective judgment, as well as objective measures, so as to increase the heterogeneity between strata and the homogeneity within strata.

6° The estimates and their sampling errors

a) The unbiased estimates

Let x_{ijk} be the test-score of the k-th pupil in a secondary sample

*⁾ The Ministry of Education annually conducts a nation-wide test in all the public lower secondary schools, and in all the general courses of the public upper secondary schools.

**⁾ The other courses—the majority of them is a fishery course—are left out because they are small in size and the survey on them is particularly difficult.

***⁾ We classify these schools into three (good, medium and poor) categories.

from the j-th school in a first stage sample in the i-th stratum.

According to our sampling procedures, a simple unbiased estimate of the population mean $\overline{X}$ is

$$\bar{x}=\frac{\sum_{i=1}^{R} N_i \bar{x}_i}{N} \tag{1}$$

where $\bar{x}_i=\sum_{k=1}^{n_i} x_{ijk}/n_i$ is the sample mean of a simple random sample of n_i units from the j-th selected school in the i-th stratum, N_i is the number of pupils in the i-th stratum, R is the number of strata, $N=\sum_{i}^{R} N_i$, and $n=\sum_{i}^{R} n_i$.

Further, the size of a sample drawn from each stratum is proportionate to the size of the stratum. Then $\bar{x}$ is reduced to a simple form

$$\bar{x}=\frac{\sum x_{ijk}}{n}\,. \tag{2}$$

b) The sampling error of the simple unbiased estimate

The variance of $\bar{x}$ is

$$\sigma_{\bar{x}}^2=\sum_{i=1}^{R}\left(\frac{N_i}{N}\right)^2 \sum_{j=1}^{M_i} \frac{N_{ij}-n_i}{N_{ij}-1}\, \frac{\sigma_{ij}^2}{n_i}\left(\frac{N_{ij}}{N_i}\right)+\sum_{i=1}^{R} \sigma_{b_i}^2\left(\frac{N_i}{N}\right)^2 \tag{3}$$

where

$$\sigma_{ij}^2=\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} (X_{ijk}-\overline{X}_{ij})^2$$

$$\sigma_{b_i}^2=\sum_{j=1}^{M_i} \frac{N_{ij}}{N_i}(\overline{X}_{ij}-\overline{X}_i)^2$$

$$\overline{X}_i=\sum_{j=1}^{M_i} \frac{N_{ij}}{N_i}\overline{X}_{ij}$$

$$\overline{X}_{ij}=\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} X_{ijk}$$

X_{ijk} : test-score of the k-th pupil in the j-th school in the i-th stratum

N_{ij} : number of pupils in the j-th school in the i-th stratum

M_i : number of schools in the i-th stratum.

We assume

$$\frac{N_{ij}-n_i}{N_{ij}-1} \doteqdot 1-\frac{n_i}{N_{ij}} \equiv \bar{a} .$$

Then we get as an approximate formula

$$(4) \qquad \sigma_{\bar{x}}^2 = \sum_i^R \left(\frac{N_i}{N}\right)^2 \frac{\bar{a}}{n_i} \sigma_{w_i}^2 + \sum_i^R \left(\frac{N_i}{N}\right)^2 \sigma_{b_i}^2$$

where

$$\sigma_{w_i}^2 = \sum_{j=1}^{M_i} \frac{N_{ij}}{N_i} \sigma_{ij}^2 .$$

Since $n_i = nN_i/N$, (4) becomes

$$(5) \qquad \sigma_{\bar{x}}^2 = \frac{\bar{a}}{n} \sigma_w^2 + \sum_{i=1}^R \left(\frac{N_i}{N}\right)^2 \sigma_{b_i}^2$$

where σ_w^2 denotes $\sum_i^R \sum_j^{M_i} \frac{N_{ij}}{N} \sigma_{ij}^2$.

c) An estimate of the variance and its components induced from other data

i) Lower secondary schools

For the purpose of calculating the value of the variance for the simple unbiased estimate and its components, the data for 500 schools, which are a random sample in the national test of 1959 and 1962, were used.

The results in 1959 and 1962 are shown in Table 2, and given for

Table 2. Sampling error and between-school variance.

Variable for stratification	Data in 1959 national test		Data in 1962 national test
	Estimated rel-variance of sampling error	Estimated between school rel-variance	Estimated between school rel-variance
α-score	0.00019	0.000076	0.000069
Type of district	0.00029	0.000178	——
Unstratified*	0.00037	0.000255	——

Rel-variance of sampling error $= \frac{1}{\bar{X}^2} S_{\bar{x}}^2$

Rel-variance of between-school component $= \frac{1}{\bar{X}^2} \sum_i^R \left(\frac{N_i}{N}\right)^2 S_{b_i}^2$

$$S_{b_i}^2 = \sum_j \frac{N_{ij}}{N_i} (\bar{x}_{ij} - \bar{x}_i)^2 \qquad \bar{x}_i = \sum_j \frac{N_{ij}}{N_i} \bar{x}_{ij}$$

* For the standard of comparison, we consider the case in which the same number of schools are selected with probability proportional to size with replacement from the population without any stratification.

the rel-variance (square of the coefficient of variation) where the number of selected schools is 200, and the number of pupils is 2,000.

As to private schools, data are not available, but the number of them is relatively small. Therefore, the type of school, the type of district and the enrollment might be enough as variables for stratification, and the value of the estimated variance is not expected to increase significantly.

Taking the above into account, we have the following scheme.

Table 3. Allocation of the numbers of selected pupils in the case of lower secondary school.

Types of lower secondary schools	Number in the population		Allocation of sample	
	Number of schools	Percent of pupil in the population total	Number of strata (≒Number of sample school)	Number of pupils
Public schools	11,861	96.0	200	2,000
Public schools for which data are insufficient	59	0.3	1	7
Private schools	535	3.7	8	77
Total	12,455	100.0	209	2,084

ii) Upper secondary schools

With regard to the general course of public upper secondary schools, we have sampled data of the national test in 1959.

In order to stratify the population and in order to get an approximate value of the variance, those data were used.

As to the general course which includes a majority of upper secondary school pupils, there is high average correlation between the mean score of the school in the test, and the percentage of pupils intending to go to college, or the number of pupils, or the type of district.

The multiple correlation coefficient among these three characteristics

Table 4. Estimated means and between-school variances.

Types of schools (Public upper secondary school)	Estimated value (data in 1959 national test)	
	$\overline{X}$ (mean)	σ_b^2 (between-school variance)
General course	43.3	158.9
Technical course	39 4	85.5
Commercial course	26.5	56.6
Agricultural course	14.2	23.6
Home economics course	14.4	20.3

Table 5. Sampling error and between-school variance (in the case of the general course of public upper secondary schools).

Variables for stratification	Data in 1959 national test			
	Estimated rel-variance of sampling error ($R=150$)		Estimated between-school rel-variance	
	A (st**)	B	C ($R=200$)	D ($R=150$)
Value from regression estimate	0.00026	0.00031	0.00012	0.00018
Ratio* and enrollment	0.00029	0.00033	0.00016	0.00020
Type of district and enrollment	0.00045	0.00049	0.00027	0.00036
Unstratified	0.00064	0.00069	0.00042	0.00056

* Ratio: See the note in Table 8, page 75.
** (st) : The case of stratifying the pupils.

Table 6. Allocation of the numbers of selected pupils in the case of upper secondary school.

Types of upper secondary schools	Number in the population		Allocation of sample	
	Number of courses	Number of pupils in the population (%)	Number of strata (≒Number of selected courses)	Number of selected pupils
General course of public schools	1,636	425,541 (*38.2*)	151	1,998
General course of private schools	791	233,142 (*20.9*)	83	1,095
Technical course of public schools	368	83,701 (*7.5*)	23	393
Technical course of private schools	130	39,115 (*3.5*)	10	184
Commercial course of public schools	568	104,818 (*9.4*)	36	492
Commercial course of private schools	342	86,083 (*7.7*)	31	404
Agricultural course of public and private schools	545	64,987 (*5.8*)	22	305
Home economics course of public schools	640	45,871 (*4.1*)	16	216
Home economics course of private schools	211	28,511 (*2.6*)	10	134
Others	19	2,782 (*0.3*)	1	13
Total	5,250	1,114,551 (*100.0*)	383	5,234

and the mean score of the school was 0.858. Therefore, we conclude that these three will be appropriate variables for stratification.

Since the average scores vary considerably from course to course, each of the courses is separately stratified.

Table 1 shows that the numbers of pupils in these courses are comparatively small and the between-school-variance within each course is comparatively small (see Table 4). So the type of district and the enrollment will serve as stratification variables for our purposes.

By using the data in 1959 sample test, we calculated the values of the estimates of the variance and the gains due to this stratification.

The results are shown in Table 5, where the numbers of strata (courses) are 150 and 200 respectively, and the number of selected pupils is 2,000. Column A in Table 5 shows the values obtained in the case of stratifying pupils by such factors as those which are highly correlated to the scores mentioned previously.

Taking this into account, we set up Table 6.

With regard to private upper secondary schools, data were not available which would have enabled us to calculate an approximate value of the variance. Then, we used the above-mentioned principles for stratification and classified schools into strata.

3. Analyses concerning the sampling design of public lower secondary schools

i) Materials used

In designing a sample for public lower secondary schools, we examined the scores of mathematics in the nation-wide achievement tests annually conducted by the Ministry of Education, on the assumption that there would be high correlation between these scores and the scores to be obtained in the I.E.A. Tests.

We investigated the following data by using a random sample of 500 schools (about 1/24 of all the schools):

a) the school means of scores (in 100 points) in the above mentioned survey for the ninth graders in the school year 1959,
b) the school means of those as above mentioned for 1961,
c) the number of pupils in the same grade in each sample school for the school year 1961,
d) the ratio of pupils intending to go to schools at higher levels to the total enrollment, for the school year 1961,
e) the features of the school district (see Appendix page 77).

The coefficient of correlation between the school means of scores for the two years came out to be 0.76, showing that high stability did not exist in the scores.

Moreover, we could not get reliable regression estimates for the school means of scores in the school year 1961 on the basis of the other characteristics that we discussed already. However, as it was probable that the I.E.A. Tests and the data cited above were closely correlated, we grouped schools by the following method which took account of all these characteristics for stratification of schools.

ii) α-score method used for stratification of schools

Let $a_k(i)$ be the value of the i-th unit on the k-th characteristic, and w_k the weight assigned to the k-th characteristic. Consider the linear sum of r characteristics in the i-th unit,

$$\alpha_i=\sum_{k=1}^{r} w_k a_k(i) \; . \tag{6}$$

In order to obtain the value of the weight w for each characteristic which gives the maximum variance of α under the condition that $\sum_k w_k^2$ is constant, we consider the Lagrangian G,

$$G=\sum_{k,j} w_k w_j C_{kj}-\lambda[\sum w_k^2 - \text{const}] \tag{7}$$

where

$$C_{kj}\equiv\sum_i \frac{a_k(i)\,a_j(i)}{N}-\bar{a}_k\bar{a}_j$$

$$\bar{a}_k=\frac{1}{N}\sum_i a_k(i)$$

and the first term of the equation (7) is σ_α^2, the variance of α. The optimum values of the w_k are obtained by taking derivatives of G with respect to the w_k and setting each of the derivatives equal to zero. Thus we have the following k simultaneous equations

$$\frac{\partial G}{\partial w_k}=\sum w_j C_{kj}-\lambda w_k=0 \; . \tag{8}$$

Let C be a $r\times r$ matrix (C_{kj}) and w a column vector $\{w_k\}$. Then we have

$$Cw=\lambda w \; . \tag{9}$$

Therefore, we solve the characteristic equation of C, and the optimum values w are obtained as the characteristic vector corresponding to the largest characteristic root of C. (In the actual calculation, we used the normalized value of $a_k(i)$.) Thus we have

$$\alpha_i = \sum w'_k a'_k(i)$$

where

$$a'_k(i) = \frac{a_k(i) - \bar{a}_k}{\sigma_{a_k}}, \qquad w'_k = w_k \sigma_{a_k}.$$

In the same manner as mentioned above, we have the characteristic equation

$$Rw' = \lambda w'$$

where R is the correlation matrix (r_{kj}), r_{kj} being the correlation between $a_k(i)$ and $a_j(i)$, and w' is the column vector $\{w'_k\}$. Then, we calculate the characteristic values and characteristic vectors of the correlation matrix R, and we obtain the characteristic vector corresponding to the largest root of R.

From the above five characteristics, we computed α_i. Table 7 shows a correlation matrix among the characteristics, and Table 8 shows the

Table 7. Correlation matrix ($\equiv R$).

	1961 Score	1959 Score	Enrollment	Ratio*	District
1961 Score	1.000000	0.755938	0.500371	0.619457	0.540323
1959 Score		1.000000	0.521824	0.555672	0.540537
Enrollment			1.000000	0.613317	0.767058
Ratio*				1.000000	0.620303
District					1.000000

Table 8. Characteristic vector and weight.

	Characteristics vector corresponding to the largest root of R ($\equiv w_k'$)	Standard deviation of a_k ($\equiv \sigma_{a_k}$)	Weight of a_k $\left(\equiv w_k = \frac{w_k'}{\sigma_{a_k}}\right)$
1961 Score	0.44702319	9.033825	0.049483
1959 Score	0.44108848	10.931606	0.040350
Enrollment	0.44606804	476.86476	0.000935
Ratio*	0.44651491	20.53047	0.022267
District	0.45525704	5.195190	0.087630
Largest root	3.4142321		

* Ratio in these tables denotes the ratio of those pupils who indend to go to schools at higher levels to the total enrollment of that grade.

characteristic roots and the corresponding vectors and weights.

As we have shown in 5° of section 2, it is better to use α_i as the variable for stratification ; because its value is much more stable than that of raw-data. Further we can see that there is high average correlation between α_i and the test score (in 1962). Table 9 shows the correlation coefficients between them and the multiple correlation coefficients among the above five raw-data with the test score in 1962.

Table 9. Correlation coefficient

	α_i	Multiple correlation coefficient using the raw-data of five characteristics
1962 test score	0.831	0.819

Therefore, we intended to calculate the α_i's of all public lower secondary schools by this method. The data used for the calculation of α_i are shown in section 2, and in computing all α_i's, an IBM 7090 was used.

iii) The number of strata

The formula (5) for the variance of the estimate shows that the first term of (5) gives the within-school contribution and depends only on the sample size n, while the second term gives the between-school contribution to the over-all variance, and depends only on stratification.

Then, to evaluate the gains due to stratification, the between-school variance contribution, the second term of the formula (5), is used. The values of rel-variance are shown in Table 10.

Table 10. Estimated value of between-school rel-variance.

Data used	Mean score $\overline{X}$	Number of strata R $R=19$	$R=38$
National test in 1959	44.4	0.000747	0.000372
		$R=20$	$R=40$
National test in 1961	15.6	0.000684	0.000330

(where the strata were formed based on grouping the α-score)

Therefore, if we take into consideration that, generally speaking, the between-school variance remains almost the same even when the number of strata, R, increases, then the second term in formula (5) is inversely proportionate to R. Thus, it may be expected that if strata are further divided into, say 200 strata, the between-school rel-variance will be a tenth of that of the sample with 20 strata. Hence, we obtain Table 2.

Acknowledgements

We are grateful to Dr. C. Hayashi (I.S.M.) and Dr. S. Kubo (Prof. of the National Institute for Educational Research) and many others for their helpful discussions and suggestions. Thanks are also due to Mr. H. Nishimura (the I.B.M. Data Center of Japan) for programming our calculations.

Appendix

the classification of districts

The Ministry of Education asked the public lower and upper secondary schools involved in the National Test to fill out the face sheet of the test of mathematics. One of the items in the sheet was the type of district as described below. We used these data in classifying public lower secondary schools.

Types of district

1) mining area in the city
2) industrial area in the city
3) commercial area in the city
4) residential area in the city
5) all other types of area in the city
6) mining town
7) fishing village
8) lumbering village, farm village in the mountains
9) local town
10) suburban farming district
11) lumbering and agricultural village
12) agricultural village with a high percentage of farmers
13) agricultural village with a low percentage of farmers
14) others

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] International Project on the EVALUATION OF EDUCATIONAL ACHIEVEMENT, UNESCO Institute for Education, Bulletin No. 2, April 1963.

[2] C. Hayashi, "Sample survey and theory of quantification," *Bulletin of I.S.I.*, 38 (1961), Tokyo.

[3] C. Hayashi, "Theory of quantification and its examples, VI," *Proc. Inst. Statist. Math.*, 9 (1961).