

# Sampling Design in Literacy Survey

Chikio Hayashi.

Fumiyuki Maruyama. Masatsugu D. Ishida.

Setsuko Takakura. Masako Taguma. Michiyo Suzuki.

(Received June 10, 1950)

The Literacy Survey was performed under the direction of the Committee organized for this purpose in August, 1948. This short report is that of sampling design in this survey, a part of works done by the statistical committees. Under some restrictions, for example, time, expenses, or testers required, this design was made.

## I. Introduction

"Literacy" is defined as the minimum reading and writing abilities and activities essential to lead a normal social life as a member of the society. The aim of this research is to acquire some knowledge about not only the literacy of the Japanese but also to find out the factors influencing the reading and writing abilities and activities. In order to answer this purpose we planned this survey. First, the basic course was discussed, and the concrete method of this survey was decided.

- (i) This survey is performed under the so-called paper "test" form.
- (ii) Papers is filled with test materials (items) necessary for diagnosing the writing and reading abilities and activities.
- (iii) These test materials (items) are prepared through the reasonable process.
- (iv) Testees are tested with these papers in the form of group test.
- (v) Testees are sampled from all over the country by the random sampling technique.

This survey consists of two important points. One is the construction of test items from the reasonable point of view, the other is the sampling method.

Here we are going to set forth this sampling method somewhat minutely as follows.

## II. The method of testing the reading and writing abilities and activities

Now, the test materials having been constructed, the testees are to be

made answer by filling in each entry of the test paper, instructions given in the same way as in a general intelligence test. In order to carry out the test as strictly as possible, the testees should be tested under the same condition. To do this, they are to be called to a certain place, where the test with its instructions is given within certain limits of time.

This test is to be given to Japanese of ages from 15 to 64 (after the Japanese fashion) who can lead normal social lives. Those who fall under the category, are counted about 45,000,000. Evidently it is impracticable, and is also insignificant when viewed from the nature of this project, to test all of them in this way. Further, testers will naturally be restricted in number, as they should be equipped with a certain ability and be trained in a certain way for this rather complicated survey.<sup>1)</sup> Moreover, the length of time required for the survey must be short, for to spend much time cannot answer the purpose of the project. Finally, the expenses also were limited. Taking these practical motives into account, we must resort to a sampling survey.

Now, if sampling is used at all, what kind of its method should be adopted? The method of random sampling is considered to be the best, for the degree of confidence can be scientifically, i. e., after the theory of probability, guaranteed by it. In our survey we adopted the stratified random sampling method by which the universe is divided into strata as homogeneous as possible from which samples are drawn at random.<sup>2)</sup>

### III. Pretest

To carry out a survey, so entirely new and of large scale as this, we must see at first, what measures should or can be taken, and which are the best ones in view of many practical restrictions. To prepare for the measures to be taken in the main test, we must make some pretests with certain designs. From the analysis of the results of such pretests, it is necessary for us to find out various eventual difficulties and to provide measures against these difficulties. Further, basic materials for the construction of the final plan of the main test can be obtained from the results. For this purpose, we effected the pretests in three places viz., in the city of Odawara<sup>3)</sup> as an example of cities, and at Nomotomura, Hiki-gun, Saitama

1) We considered well trained teachers of primary schools at each test spot as appropriate testers.

2) Since such sampling method is taken, the surveyors have to be trained in the techniques of sampling. It is not conceivable that the training may impose so much burden upon them, if it is given to them together with the drill of test instructions.

3) In Odawara, the pretest was made in two days. On the second day, those who could not come for inevitable reasons on the first day, attended the test.

Prefecture, and at Nagaura-mura, Kimitsu-gun, Chiba Prefecture, as two examples of rural areas. Thereby we acquired experiences which were useful for the administration of the main test.

From the analysis of the results we got some important informations for the testing and sampling which were utilized for drawing up the final plan :

- ( i ) Considerable revisions were brought into the test materials.
- ( ii ) The general method of testing and sampling was checked.
  - (A) The test instructions were revised.
  - (B) The following facts were ascertained and consequently utilized for the final sampling design :
    - (a) The difference of urban and rural areas in the reading and writing abilities.
    - (b) The kind of industry affects the reading and writing abilities.
    - (c) The habit of reading newspapers affect a great deal the reading and writing abilities. Among those who lead lives in the same area and have the same kind of industry, those who have the habit of reading newspapers have higher abilities.
- (iii) The method of research on cultural backgrounds of testees was determined—the forms of the data sheet and the testee list were established.
  - (A) The data sheet is to be entered by each testee at the testing day, and serves as indices of his cultural background, being provided with columns for the date moved to, age, sex, industry, occupations, the year of employment, the former industry and occupation, industry and occupation of the family, school experiences, reading of newspaper, listening of radio, the data and hour of arrival, etc.
  - (B) The testee list is to be entered the names of testee and serves for checking testees at the testing day.
- (iv) The sampling method to be actually taken was examined. In general, the stratified random subsampling method was ascertained to be best.

At the last stage of sampling, that is to say, the sampling of individuals, the systematic sampling method from the rationing records was decided to be adopted. Then the sampling interval must be carefully decided not to synchronize with the average

number of the members in a household.

- (v) The way of calling testees to the testing places was examined. It was decided that the letter to call was to be handed over to each testee through the local administrative officials and the teachers of primary schools of each test spot.
- (vi) The attendance of testees was observed. Then to 18 percent of testees did not attend the test for inevitable reasons: disease, birth, official travel, etc. However, the same distribution in regard to sex and age was seen both in presentees and in absentees.<sup>4)</sup> The score distribution and the mean value of test scores of those who were absent on the first day for inevitable reasons at the Odawara pretest, did not show the significant difference with that of the presentees.<sup>5)</sup> From these results it may be assumed that the existence of absentees for inevitable reasons will not give any bias to the whole results. Further it was required for the main test that necessary measures should be taken lest the absentees should be more than 20 percent of the sample.

4)	Number		Number			
	Sex	Presentees	Absentees	Age	Presentees	Absentees
	Male	202	42	15-24	134	22
	Female	208	49	25-34	89	26
				35-44	85	17
				45-54	60	12
				55-64	42	14
	Total	410	91	Total	410	91

The significant test of these results is as follows:

	Value of $\chi^2$	Degree of Freedom	Probability for it
Sex	0.29	1	0.7-0.8
Age	5.28	4	0.2-0.3

5)	Day	
	First	Second
Score		
0-65	177	12
66-70	93	6
71-75	109	13
Average	59	63
Standard Deviation	19	16

The significant test of this:

	Value of $\chi^2$	Degree of Freedom	Probability for it
Score	1.58	2	0.3-0.5

#### IV. Sampling Design

##### (a) Stratification

Next, making use of the results obtained from the pretests, we established a sampling design and a concret plan for the survey. It is rather difficult to draw each individual with equal probability in such a large scale sampling. If we use a conventional method as is usually done, it may cause biases. To draw up the design strictly, it is very useful to divide the universe into homogeneous groups. This process is called stratification, and the number of elements belonging to each group, i. e., stratum, is required to be made small reasonably. In our sampling we followed this stratification and decided to make those groups which are considered to be homogeneous in terms of reading and writing abilities.

Before this we must decide the sampling unit and the sampling system. Taking various conditions into account, sub-sampling system will be adopted and primary sampling units are decided to be wards, cities and districts (Gun). Secondary sampling units are villages and towns or individuals, tertiary sampling units are individuals.

Population is constructed from the universe which consists of the Japanese from 15 to 64 who live in Japan except some remote spots, the ratio of the population size of which is 0.673% to the total. What are the sampling probabilities of various sampling units?

Primary sampling units and 2nd sampling units (villages and towns) are drawn with the probabilities proportional to the population size. Secondary and tertiary sampling units (individuals) are sampled with equal probabilities from the sampling units to which they belong. Thus population has been constructed, we go a next step.

Now we consider the stratification of the primary sampling units.

For this purpose, we have to consider the factors affecting the abilities. The factors which are taken into consideration for the stratification are naturally assumed and acquired by the pretest. These factors are reflected in the degree of diffusion and spread of culture of the residence (with respect to the primary units), that is to say, urban and rural, geographical-cultural characteristics etc.

The factors adopted are as follows.

- (A) Japan Area: Past social survey experience indicates that certain areas, rather than prefectures, may be considered critical of geographical-cultural distinctions. For the delimitation we divided the whole territory of Japan into the following six districts.

- (1) Hokkaido
  - (2) Tohoku
  - (3) Kanto
  - (4) Kansai
  - (5) Chugoku and Shikoku
  - (6) Kyushu
- (B) Urban and Rural Area: This reflects such characteristics as population size and density, economic-ecological structure and degree of urbanization. This distinction is made concrete by that between city and Gun.
- (C) Economic (industrial) type  
The types of economic activities were classified into tree.
- (1) Farming, including, (a) farming par excellence, (b) forestry, and (c) fishing.
  - (2) Industrial, (a) common industries, (b) construction works, (c) mining, (d) enterprises concerning gas and electricity, etc.
  - (3) Commercial, (a) commerce, (b) free occupations, (c) official business, and (d) other services.

These types reflect the degree of urbanization according to the economic theory of Dr. Colin Clark, so influence the abilities and activities of the people who lead social lives there, I think. In some cases the qualitative structure of industry is used.

- (D) Population Size.  
This label was used in the stratification of urban, because it is confirmed to represent the degree of urbanization by the several surveys.
- (E) The ratio of radio listeners.  
Also it reflects the degree. Especially in this case the materials of 1940 was used.  
The reason why the ratio of radio listeners was preferred is that it shows well a high correlation with the ratio of newspaper readers and other cultural factors. However, the statistics of 1946 is improper for our purpose, for listeners have increased during and after the war so that their ratio has become nearly alike to that of urban areas. Hence, the statistics of 1940 were used in our design. The details are as followings.
- (i) It is recognized that the ratio of newspaper readers influences the abilities. So it is advantageous to use this label.

But these materials were in lack with respect to the primary sampling units. We had to look for the materials related to it. This is the ratio of radio listners. The correlation coefficient between the ratio of radio listners in 1940 and that of newspaper readers when a prefecture is considered to be a unit is 0.753, while the correlation coefficient between that in 1947 and the ratio of newspaper readers is 0.456. This difference is significant. The ratio in 1940 is more representative of that.

- (ii) The ratio in 1947 has many errors because of many causes.
- (iii) Reading and writing abilities are influenced by the historical features of the residence. That in 1940 reflects naturally a index of the historical, cultural and social features of the units, as it says, but that in 1947 does not, because the Pacific War has changed forcibly the ratio of radio listners under the pressure of necessity.
- (iv) That in 1940 correlates highly with the population density. coefficient is 0.673 when a Gun is considered to be a unit.
- (v) The force of influence of urban on rural represented by the population of Gun and the adjacent cities, correlates highly with that in 1940. The coefficient is 0.681 when a Gun is considered to be a unit.
- (vi) That in 1940 correlates highly with the degree of communications.
- (vii) That in 1940 correlates highly with national income in 1938 which reflects naturally the degree of culture. The coefficient is 0.834 when a Gun or city is considered to be a unit.

Summarizing these, that in 1940 is considered to be a representative index of historical, cultural and social background of a unit.

(F) The number of higher schools.

This label is used for the stratification of some cities.

(G) Geographical-climatic conditions.

With these characteristics and others<sup>6)</sup> in view, we established the

---

6) The statistic materials which were used for the sampling design are as follows:

- i. The statistics of the total population of the Statistic Bureau, Oct. 1947.
- ii. The statistics of the population by sex of the Statistic Bureau, Oct. 1947.
- iii. The statistics of the population by industries of the Statistic Bureau, Oct. 1947.
- iv. The statistics of the population density by E. S. S., Oct. 1947.

stratification. . Thus we classified cities into four groups, namely, (1) the six big cities (Tokyo, Osaka, Kyoto, Nagoya, Yokohama, and Kobe), (2) the cities of the population more than 2,000,000, (3) the cities of 1,000,000 to 2,000,000, and (4) the cities below 100,000.

In the big cities we stratified them into ward groups according to the economic types and the industrial structure. The strata of other cities were constructed also by the industrial criteria, sometimes with the ratio of radio listeners and a certain topographical conditions.

In the rural areas, the criteria used were in the following order: (a) the degree of commercialization, four grades being set up, (b) the degree of industrialization, three grades being classified; (c) the types of the special industries, viz., fishing, mining, and heavy industries, (d) the ratio of radio listeners, and (e) either the ratio of newspaper readers, or population density, or some topographical conditions.

Considering such criteria, the order of sampling variance, the balance of so-called between and within variance and the methods of analysis, we constructed 87 strata in urban area and 76 in rural.

Then from these sampling units individuals, the ultimate samples were drawn. To make equal the probability with which each individual was drawn, as stated above, we made the probability of sampling cities, wards, etc. proportionate to the population contained in them. We explain the stratification of secondary sampling units. Each primary sampling unit drawn from each stratum was further substratified as representative of the unit as possible, that is to say, as homogenous villages and towns in the unit as possible were grouped into substrata. Making use of various kinds of maps we made the substratification chiefly in regard to the conditions of railways, population density, land utilization, the kind of industry etc. After the substratification, only one town and village was drawn with probability proportionate to the corresponding population size from each substratum.

Individuals were drawn from wards, cities, towns and villages. At the sampling of individuals, following the method of the systematic sampling with random strat, we drew them from the rationing records which are best kept in local administrative officials. From the results of the pretests, special care was taken in using the records. We laid down the principle

- v. The ratio of radio listeners in the Business Report Degect of Japan Broadcasting Society, Dec. 1940 and Dec. 1946.
- vi. The ratio of newspaper readers, the Paper Allocation Section, Nov. 1947.
- vii. The number of Schools, the Education Ministry, April 1947.
- viii. The income of districts, the Finance Bureau of Tokyo Metropolis, 1938.



that the number of the sample interval should be a prime number so that it may not be a multiple of the average number of family numbers.

For the execution of sampling it was decided that the sampling was to be done up to the drawing of towns and villages by the hand of the members of the PAC grounds and that the sampling of individuals was to be entrusted to the local testers who were trained beforehand.<sup>7)</sup>

### (b) Determination of Sample Size and Allocation of Samples

We assume that the reading and writing abilities and activities are expressed by the scores. These scores are the labels of the elements of our population. So we will find the distribution of scores in it. Let us estimate the distribution under the desired degree of confidence. First, for the simplicity, suppose that  $n$  samples are drawn from the population of size  $N$  which is constructed by giving the  $N$  elements equal sampling probabilities, and the distribution is estimated. Let  $A_1, A_2, \dots, A_i$  be the labels of scores in the population,  $p_1, p_2, \dots, p_i$  be the distribution of the labels respectively. Of course  $\sum_{i=1}^i p_i = 1, p_i \geq 0$ . In order to estimate the distribution at the desired degree of confidence, sample size  $n$  is determined so that the following formula may hold.

$$\frac{N-n}{N-1} \cdot \frac{p(1-p)}{np} \leq \varepsilon^2 \quad p = \min(p_1, \dots, p_i),$$

where  $\varepsilon$  is a constant to determine the degree of confidence. It is easily understood that the coefficient of variations for the estimates of  $p_1, \dots, p_i$  are smaller as this method is used.

Suppose the number of testees elements of population with a certain label is counted 5 percent in the distribution. The size of samples can

---

7) As the ultimate samples are drawn from cities, wards, etc. by the random sampling, an error committed in this procedure will affect greatly the whole results. However, precisely the design may be made, if the instructions of sampling happen to be neglected at any test spot, the survey will doubtless lose its value. Accordingly, we explained the details of method in the Instruction Guide. Here we paid special attention to the meaning of the word "random." By "random" one is commonly apt to understand optionality or arbitrariness. It is imaginable that one may draw samples as one pleases. If the sampling is done in this way, a certain systematic tendency may appear in the whole result instead, and consequently the hypothesis on the random sampling will lose its ground. It should be kept in mind that the expression "at random" finds its meaning only in the manner of how the whole samples are drawn. Therefore, in order to be able to perform the random sampling, the word should not be used in the instructions to the local surveyers. Thus, we planned to train the testers to do just as the Instruction Guide explains, in which the table with the sample size allocated to each spot, start number, sampling interval, is given.

be determined so that the estimated value from the samples in regard to such a label may have the precision  $5\% \pm 0.5\%$  in terms of the confidence interval, if the degree of confidence is desired to be 99.7 percent. If the samples can show this precision at 5 percent, it can be understood that the coefficient of variation will be smaller at the distribution of more than 5 percent. Thus, we obtained 17,100 necessary samples. This value is determined from the above formula where  $\varepsilon = \left(\frac{0.005}{3}\right)^2$ . With so many samples, a considerably minute analysis is possible and the sample size at each test spot may not be too small to have any significance.

But we adopt the sub-sampling system. So the degree of confidence may be diminished. This order must be appreciated. Like the above mentioned, consider the label corresponding to 5%. Every element which has this label has the attribute 1, which has not had the attribute 0. Let  $\bar{x}$  be the estimated ratio of the elements which have the label corresponding to 5%, where  $\bar{x}$  is the unbiased estimate from the samples of the population value. Then the variance of the estimate is proved easily to be approximately as below

$$\sigma_{\bar{x}}^2 \leq \sum_{i=1}^R k_i^2 \sigma_{b_i}^2 + \frac{\sigma^2}{n}$$

where  $\sigma_{b_i}^2$  is between variance in the  $i$  stratum,  
 $k_i$  is of the population size the ratio of the  $i$  stratum to the total,  
 $n$  is sample size,  
 $R$  is the number of stratum.

The first term may be appreciated to be 0.0000030 and the second term to be 0.0000028 as the result of estimating the between variances from the various points of view in urban and rural strata. So the variance of  $\bar{x}$  is considered to be approximately twice the order of the variance in the simple case. Though sub-sampling method is adopted in some strata, the diminution of the degree of confidence is considered to be negligible, following the same consideration processes as the above mentioned. And such an estimation of variance seem to be over-estimated because we neglect in some sense the effect of stratification. As result the size of sample 17,100 will be sufficient for this survey.

Having fixed the sample size in this manner, we proceed to allocate them. This sampling design, however, is intended for estimating the distribution of population in terms of scores. An allocation such as is optimum for a single label is considered to be improper in the survey consideration which is so complicated that the reading and writing abilities

can be measured only by many criteria combined. Besides the reasearch also required to get statistic materials about schooling, sex, age, occupation, listening of radio, reading of newspaper, years of dwelling, conditions of assembling, as well as the cross tabulation of these statistics. Further, the counting is not merely restricted to the estimation of the average value or separate indices. A minute analysis is also to be done. Therefore, in view of the conveiences of the counting, the samples were allocated in proportion to the population in each stratum.

However, from the results of the pretests, we found that the ratio of inevitable absentees in about 20 percent in urban and about 10 percent in rural. Including non-responses which may not give any bias to the whole, the sample size was determined to be 21,008.

### Complements

#### (a)

The main test of this survey was performed in August of 1948. The inspectors were sent to the test spots selected at random to check whether the plan was performed pursuant to the instruction at the test spots or not.

After the main test, pretests were performed to correct the results, in the spots, the performance at which was discovered to be improper, and the spots, the ratios of absentees at which were low.

And so the validity test was performed to see whether the test items has validity and reliability or not. Then the school test was to compare the results of the main test with those of the students to high schools from elementary schools.<sup>8)</sup>

#### (b)

The effect of the stratification was calculated after the main test. Suppose that the population mean of the scores will be estimated. The within variance was easily estimated. The between variance was reasonably estimated from the relations between the values estimated from the hyper-regression plane—the statistical expression of the numerical relations between the scores and the characteristics for stratification of 1st sampling units—and the actual results. As this result, the effect i. e.  $\frac{\sigma_{\text{non-st}} - \sigma_{\text{st}}}{\sigma_{\text{non-st}}}$  where  $\sigma_{\text{non-st}}$ ,  $\sigma_{\text{st}}$  are the sampling variance of the sample mean respectively in the case of non-stratification and of stratification, turned out to be about 50%. This stratification will be said to be successful.

*The Institute of Statistical Mathematics*

8) The details of this survey will be in press in Japanese, "Leteracy of the Japaneses" 1950 Tokyo, Todai Kyodokumiai Shuppanbu.