The Development of Non-Language Test of Verbal Intelligence

S. CHATTERJI & MANJULA MUKERJEE
Indian Statistical Institute, Calcutta

Introduction

Ability to profit from academic training is frequently identified as intelligence, and tests known by this name are generally validated against educational proficiency attained. Hence it would be quite justified to identify intelligence with scholastic aptitude and consider it as the indicator of the potential ability of the student to succeed in school or college.

It is a general practice to use a scholastic aptitude test along with two or three achievement tests for the purpose of selecting students for admission into various training courses. There have been a number of studies investigating the relation between aptitude test scores and achievement in the west and "the correlation between intelligence tests and composite standing of the pupils may be said.....to lie usually between .40 and .60. Probably, in the majority of the cases the correlation will be found to be in the neighbourhood of .50 but under very favourable conditions it may be somewhat above this" (3). These relationships are high enough to be useful in predicting academic success. Few studies done in India seem to suggest that the situation is not very different here. There are verbal tests of intelligence available in several Indian languages. However, western experience has indicated the value of having both a verbal and a non-verbal test on which comparable norms have been developed. The Weschler-Bellevue & Large-Throndike may be cited as familar examples. Such a non-verbal instrument could be profitably used to determine whether a low score in the verbal test is due to lack of ability or due to specific reading handicap. Another use may be with children whose linguistic background may differ somewhat from that of the children on whom the verbal test was standardized. According to Doppelt (2), such a culture-fair test should lead to better prediction because it would pick out the individual cases who are trainable despite poor cultural background. But it is necessary to note that the index of ability as obtained through performance tests is not as stable as one would desire and the usual type of non-verbal tests is less efficient in predicting the scholastic success than the available verbal measures.

These problems were discussed at the All India Workshop for Long Range Planning in the Testing Field held at Bhopal and it was suggested (5), that the solution was to develop a test which would measure the verbal ability with non-linguistic or language-fair material. The items for this test should be developed in

Manuscript received on February 14, 1966.

such a way that they would require words to solve them but would not specify the language from which the words are to be chosen and used. Non-verbal symbols, Abstract symbols, Concrete symbols (pictures and numerical symbols) can be used. The materials would all be culture-fair *i.e.*, common to all Indian sub-culture and non-linguistically presented. Answer is also to be indicated without the use of language. Language would be only the assumed intervening variable, between the visual presentation and non-verbal answer marking ; it would be something which goes on in the head of the candidate using any verbal symbol which he wishes to use. The items included in this type of non-language test would be similar to the one generally used in verbal reasoning tests ; the only difference being in the fact that in place of language, pictures would be used. Such a test would be quite different from the usual non-verbal tests which use both pictures and diagrams as the medium. Hence to distinguish this new type of test from the non-verbal tests the term " non-language " has been coined.

By the way, it might be mentioned here that the present state of linguistic transition in India makes the use of non-language or language-fair tests the potentially fairest method for doing this. Moreover in a country like India with so many regional languages it is highly important to develop parallel tests in several Indian languages for providing an All-India standard. In order that the several versions of the same test are equated on a comparable scale it is necessary to have a common test administered, on the basis of which the various versions of the same test in different languages could be compared. Non-language tests would serve well in this respect.

It is better to consider the item types which can be included in such a test. Items like Anologies, Classification, Similarities, Opposites, Picture Series etc., which are generally included in a Verbal Reasoning Test can also be used here. A brief description of these types are presented below :

(A) *Analogy* :—In this type there are three pictures and a blank space in the item stem and these are followed by four suggested answers (pictures). There is a specific relation between the objects presented in the first pair and the task is to find out from among the suggested answers that picture which would fill up the blank space in such a way that the second pair also has the same relation between them.

(B) *Classification* :—In this type, five pictures are presented one of which is different from the rest in some respect. The task is to find out this picture.

(C) *Similarities* :—Here also five pictures are presented. The candidate is required to find from among the last four pictures that one which is similar to the first one.

(D) *Opposites* :—Here instead of finding the similar one as in the preceding type the candidate is required to find out that one which is opposite to the first one.

(E) *Picture Arrangement* :—In this type three, four or more pictures are presented in each question. The task is to find out the correct order of the series of pictures. For example, four or five pictures presenting a story (as is done in cartoon pictures) are disarranged and the examinee is required to identify the correct order in which these pictures should be arranged.

(F) *Series Completion* :—Here a series of pictures is presented and it requires the candidate to choose the right one from the suggested alternatives, which would complete the series. This may be varied by leaving out the first, the second or the third, etc., rather than the last picture only.

It may not be necessary to use pictures only. While simple geometric problems may often be solved at a perceptual level, the complex ones may require verbalization for their solution. It is also possible to partial out such factors by choosing items which have a high correlation with a Verbal Test.

Level of the proposed Non-Language Test

Before constructing the test, it was necessary to decide the level for which the proposed test was to be constructed. According to the present educational system of our country, the crucial moment in a student's career comes at the class VIII level. After Class VIII, each student is allocated to one of the seven streams or courses of training. So at this point it also becomes important to assess the intelligence of each student and to decide whether further schooling would be effective or not, *i.e.*, to find out whether the student has enough scholastic aptitude to profit from further training. Hence it was decided to keep the level of the test at the Class VIII standard.

Construction of the Non-Language Test

It is better to mention at this point that the purpose of this study was to find out whether it was possible to construct a test of the type mentioned above. Hence it was planned to construct a small test with 50 or 60 items with which a pilot study could be conducted.

On the basis of what has been said items were first written using language, and with the help of an artist these verbal items were translated into pictures. The picture items were then scrutinized and modified, and by assembling these items a tentative form was constructed for a pilot study.

There were three parts in this test, containing items of the following types: (A) Similarities, (B) Classification, (C) Picture Series, in these three parts there were 23, 15 and 7 items respectively. It was also decided that the test was to be administered under power condition, so that it would be possible to do the item analysis for all the items in the test.

Two other tests were used along with this Non-Language test *viz.*, Test of Abstract Reasoning and Test of Verbal Reasoning. These two tests were earlier constructed for the class-VIII level. Both the tests consisted of 35 items each, and the time limits were 22 and 20 minutes respectively. The idea was to conduct three different item analyses of the non-language test,—on the basis of (i) the total score on the Non-Language test itself, (ii) total score on the Verbal

---

Note :—The idea of a Non-Language Test was discussed by Dr. A. Edwin Harper Jr., at the All India Workshop for Long Range Planning in the Testing Field held at Bhopal.

Reasoning test and (iii) total score on the Abstract Reasoning test. On the basis of the information obtained by such analyses, it would be possible to retain those items of the Non-Language test which would show fair discrimination values on the basis of the first two analyses and a low discrimination value from the third analysis. Then the resulting test would be internally consistent and would measure the verbal factor through a non-language medium and at the same time it would not be related to abstract reasoning as measured by the non-verbal test (*i.e.*, Test of Abstract Reasoning).

Collection of the data

After the test was constructed and printed, it was necessary to work out a sampling design and collect the data required for item analyses, etc.

It should be mentioned here that this pilot study aimed at finding out whether it was possible to develop such a test or not. Hence, at this stage it was not necessary to collect the data on a large scale. Therefore, it was planned to consider the schools at Calcutta only and draw samples from them. If it was actually possible to develop such a test then the pilot study was expected to give adequate information which would be of great help in planning and constructing the final tests in a much better way for further study on a wider scale. But at the same time it was necessary that the sample used in the pilot study should be representative of the Class VIII students of Calcutta.

From an earliar survey made by the authors certain information about the higher secondary schools is available. On the basis of the information, the Higher Secondary Schools at Calcutta were divided in to six groups as follows :

(*a*) Boys' School where more than 75% of the students sent up for the Higher Secondary examination passed the examination in 1963.

(*b*) Boys' School where less than 75% of the students sent up for the Higher Secondary examination passed the examination in 1963.

(*c*) Girls' School where more than 75% of the students sent up for the Higher Secondary examination passed the examination in 1963.

(*d*) Girls' School where less than 75% of the students sent up for the Higher Secondary examination passed the examination in 1962.

(*e*) Boys' School recently upgraded to the Higher Secondary level.

(*f*) Girls' School recently upgraded to the Higher Secondary level.

Each of these six categories contain near about 15 schools, except category ' e ' where there were forty schools. Only the schools which had Bengali as the medium of instruction were considered in this study as one of the tests *viz.*, Verbal Reasoning was in Bengali.

For the six categories mentioned above 13 schools were selected, two from each of the five categories and three from category ' c ' as there were larger number

of schools in this category. These thirteen schools were selected with probability proportional to the number of students in the schools concerned.

In each of these schools the three tests *viz.*, the Non-Language Test of Verbal Intelligence, Test of Abstract Reasoning, and Test of Verbal Reasoning were administered on all the students reading in Class VIII. The total number of students thus tested was 1,075.

Analysis of the data

After the administration of the tests, the papers were scored. As the tests were of the multiple-choice type there was some chance of guessing on part of the candidate and so instead of taking the right score only, the scores were corrected for guessing.

The means, standard deviations, modes, medians, etc., of the scores were obtained for the three tests used, and these values are presented in Table 1 along with the maximum possible scores that an examinee could obtain in each of these tests.

TABLE 1

Showing the means, modes, medians, standard deviations and the maximum possible scores of the three tests used

N = 1079	NL	AR	VR
Mean	29.47	15.60	20.77
Median	29.80	14.00	20.25
Mode	33.20	11.00	19.00
Standard Deviation	5.04	6.14	5.84
Maximum Possible Score	55	35	35

From Table 1 it can be seen that for the Non-Language test the standard deviation was not high enough though the mean score is not low. As this test was administered under a power condition this sort of outcome is quite expected. It should however be noted that though the Maximum Score is 55, yet the highest score that could be expected from this group is only 45 as Mean+ 30 is equal to 44.59 ($29.47+3\times5.04$). This suggests that the test has a wide enough range to use with higher classes as well.

Reliability and Intercorrelations among the three tests used

The product-moment correlation coefficients between the Non-Language Test and the other two tests were calculated and these coefficients are presented in Table 2 along with the correlation between these two other tests also. These correlations were calculated on the basis of 200 cases selected by using the method

of systematic random sampling from the total of 1,079 cases. The odd-even reliabilities of these tests were calculated and the obtained values (after Spearman-Brown Correction) are presented in Table 2.

TABLE 2

Showing the intercorrelations among the three tests, reliabilities and the maximum possible score

N = 200	Non-Language Test	Abstract Reasoning	Verbal Reasoning
Non-Language	...	.53	.60
Abstract Reasoning	.53	...	.48
Verbal Reasoning	.60	.48	...
Reliability	.82	.85	.92
Max. Poss. Score	55	35	35

Table 2 shows that the correlation between the Non-Language test and the Verbal Reasoning was higher than that between the Non-Language test and Abstract Reasoning. This is what is wanted, but if this relation between Abstract Reasoning and the Non-Language test was less than what was actually obtained it would have been more desirable. It was planned that while revising the test (after item analysis), those items which showed good discrimination values on the basis of Abstract Reasoning could be eliminated and it is expected that the correlation between Abstract Reasoning and the revised Non-Language Test could be reduced.

It can also be seen from Table 2 that the obtained reliability of the Non-Language test was fairly high and comparable to those of the other two tests. Moreover, it is expected that after purifying the test this value can be further increased.

Item Analysis

The most important step in the construction of the Non-Language Test as discussed earlier was to conduct on item analysis. It was planned to retain those items which were related to the verbal factor and reject those which were related to abstract reasoning. Hence item analysis of the Non-Language test was done thrice as mentioned earlier :—

(i) on the basis of the total scores obtained on the Abstract Reasoning test,

(ii) on the basis of the total scores obtained on the Verbal Reasoning test, and

(iii) on the basis of the total scores obtained on the Non-Language test itself.

First, 370 answer sheets were taken out from the total pool by using a systematic sampling method. Item analysis values were obtained by using the

method developed by Davis (1) and modified by Harper (3). The three sets of item analysis values are presented in Table 3.

TABLE 3

Showing the Davis Difficulty and the Discrimination values for the Non-Language Test on the basis of the total score on the Non-Language test (NL), the Verbal Reasoning test (VR) and the Abstract Reasoning test (AR)

Item No.	On the basis of NL Scores		On the basis of AR Scores		On the basis of VR Scores		Item No	On the basis of NL Scores		On the basis of AR Scores		On the basis of VR Scores	
	Dif.	Disc.	Dif.	Disc.	Dif.	Disc.		Dif.	Disc.	Dif.	Disc.	Dif.	Disc.
1	53	22	53	0	50	8	29	55	17	58	5	57	10
2	58	53	59	12	58	30	30	38	5	39	5	39	4
3	18	3	20	0	18	3	31	60	38	65	20	62	22
4	56	35	55	8	53	22	32	56	40	56	16	55	21
5	39	6	40	5	39	3	33	42	26	44	13	43	23
6	61	35	58	10	61	21	34	41	4	46	0	44	1
7	71	40	76	27	75	25	35	26	3	30	5	30	8
8	67	44	67	19	67	28	36	47	9	46	2	48	4
9	49	50	50	19	50	31	37	71	28	73	10	69	21
10	65	29	67	5	64	17	38	66	27	74	17	67	18
11	74	31	76	15	75	15	39	70	32	70	29	70	25
12	50	38	49	10	49	22	40	41	29	62	4	69	16
13	46	25	43	8	45	15	41	76	23	77	9	76	15
14	51	13	53	8	52	5	42	65	37	65	16	63	15
15	54	29	55	12	55	14	43	33	3	33	2	30	0
16	64	50	67	20	66	31	44	77	36	78	4	78	25
17	40	19	39	13	40	10	45	27	12	25	4	26	4
18	64	39	63	17	63	19	46	46	29	47	11	48	16
19	60	35	60	8	59	19	47	40	4	42	9	41	4
20*			Omitted				48	69	34	69	9	67	21
21	35	19	37	4	35	13	49	44	17	43	8	44	10
22	41	19	39	10	39	12	50	39	24	38	15	40	17
23	66	38	67	20	65	20	51	52	31	52	14	51	16
24	52	2	54	0	52	0	52	54	39	55	18	54	19
25	24	9	22	10	24	11	53	28	29	28	11	29	11
26	42	8	41	7	42	6	54	42	26	44	10	44	17
27	63	20	67	8	65	15	55	18	40	19	9	20	20
28	73	44	79	28	77	35	56	18	40	16	17	20	12

* Omitted as found defective.

Examination of the values obtained showed that the difficulty values of the items did not vary widely with respect to the three sets of item analysis, and theoretically the difficulty values should not vary at all. They are afterall, estimates of the true value (*i.e.*, the value for the entire sample) based on two sub-samples (*i.e.*, the 27% sample). So variation of the results from the

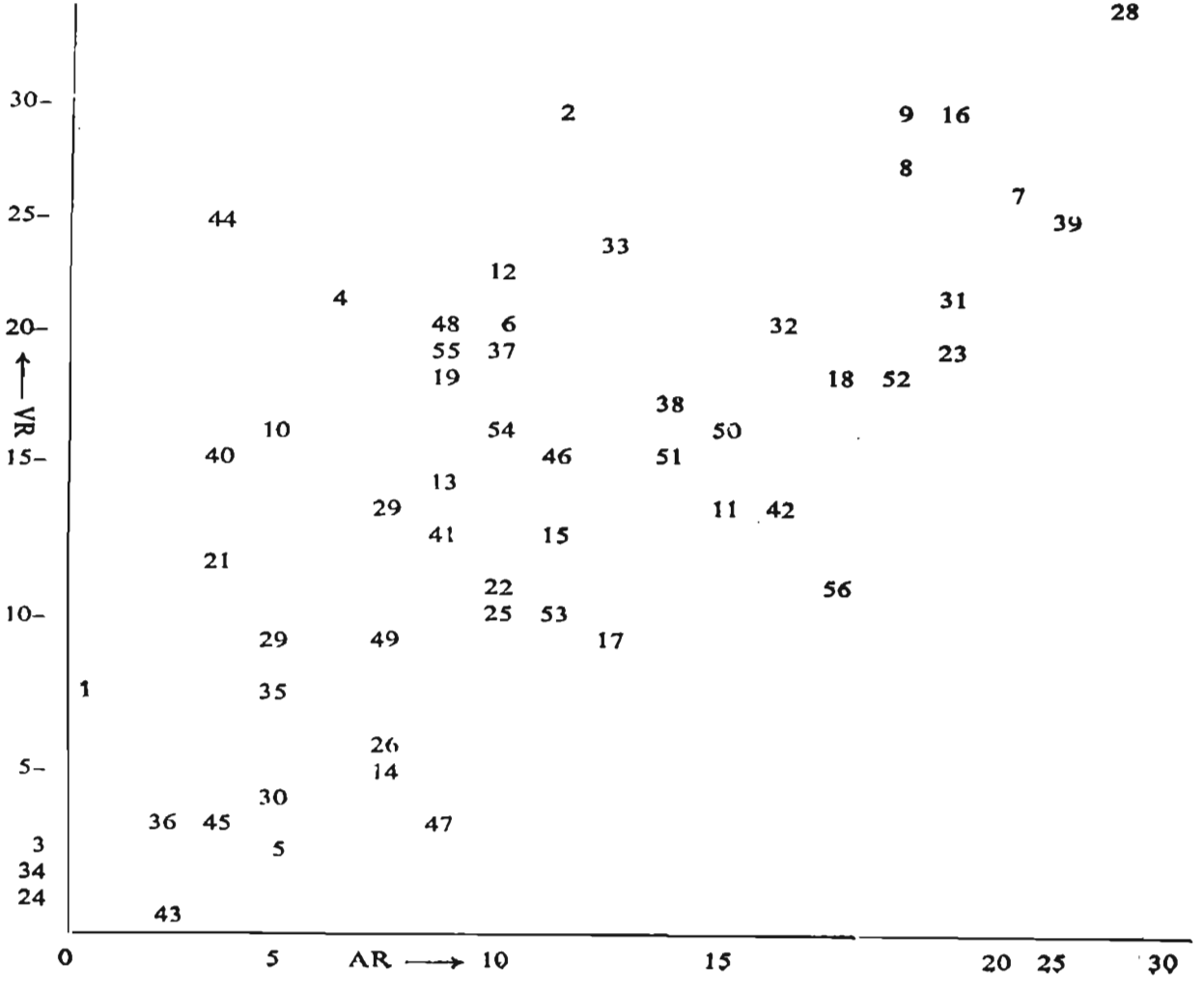
different samples should be within sampling error only. It looks from the obtained data as though this is true.

But at the same time it was found that there were wide variations in the discrimination values. It should also be noted that there are items which are related to the verbal Reasoning very well but not with the Abstract Reasoning and such items are desirable and should be retained in the revised version of the Non-Language test.

In order to select items which would have maximum correlation with Verbal Reasoning combined with minimum correlation with Abstract Reasoning a scatter diagram of the item discrimination values was plotted. This diagram is presented in Fig. 1.

FIG. 1

Showing the scatter gram for the item discrimination values of NL items based on AR and VR



The discrimination values obtained on the basis of AR scores were taken on the X axis and those on the basis of VR scores were taken on the Y axis. Instead putting dots on the scatter diagram item numbers were written to help easy identification of the items. It was decided that only those items were to be selected which had a discrimination value equal to or more than 15 on the basis of VR score and had a discrimination value below 15 on the basis of AR score. But if the difference between these two values was less than five then that item was to be rejected.

In this way 17 items were selected, and all these items also had high discrimination value obtained on the basis of NL scores. In addition to this 5 more items were also selected as their discrimination values on the basis of VR score were much high than those on the AR scores. This gave a total of 22 items for further work.

The results obtained in the pilot study may be summed up as follows :— (*a*) The standard of the Non-Language test was quite in keeping with the level of the group *i.e.*, the difficulty values of the items and the mean score of the test showed that the test was not a difficult one for this group of students. (*b*) The test was internally consistent, as near about 40 out of 55 items of the test showed fairly high (> 20) discrimination values. (*c*) There were 22 items which were suitable for being included in the revised version of the test on the basis of the criteria discussed earlier.

Validity

To start with, the NL answer sheets were first scored on the basis of the 22 selected items and this new score was to be used for validity study. This was done because these 22 items represented the type of test that was sought to be developed and this part of the study was expected to throw some light as to what could be obtained from such test.

It was mentioned earlier the NL test was administered on about eleven hundred students reading in class VIII of 13 Higher Secondary Bengali medium Schools at Calcutta. The Half-Yearly and annual examination marks for these students were obtained. As the annual examination is more important from the point of view of the students as well as that of the school, it was planned to obtain the product-moment correlation between the NL test scores, and annual examination marks for the validity study. The correlation between the annual and the half-yearly examination was also to be calculated and the obtained index was expected to give an idea as to the reliability of the criterion itself.

Table 4 shows the results obtained from different schools.

From Table 4 it can be seen that there were wide variations in the obtained mean values of the test from school to school. In one school it was as small as 12.39 whereas in another it was as high as 17.83. For the latter school the

TABLE 4

Showing the means, σ, (based on 22 items), the correlations between NL and Annual, and between Annual and Half-Yearly σ different schools

	School	No. of cases	Mean of NL score	Mean of NL score	Correlation of between NL and Annual	Correlation of between Half-Yearly & Annual
1.		131	13.26	3.67	.33	.80
2.		162	12.39	3.59	.28	.85
3.		86	15.36	3.31	.39	.76
4.		85	15.29	3.03	.36	.80
5.		23	12.78	4.09	.19	.72
6.		98	13.71	4.08	.12	.79
7.		68	17.83	2.79	.34	.77
8.		40	14.04	4.05	.16	.61
9.		95	13.98	3.71	.19	.84
10.		84	14.22	3.77	.24	.85
11.		68	13.81	3.67	.14	.88
12.		38	12.68	3.52	.27	.86
13.		50	16.35	2.67	.35	.74

test was extremely easy as the maximum possible score was only 22. The small standard deviation also shows that the test was too easy for this group. For other schools the test were not so easy.

As regards the validity coefficients, there were also vide variations among the schools. The values ranged from .12 to .39.

The maximum validity coefficient obtained was .39 and this is very high and also very encouraging *for a single test with only 22 items.* This proves the possibility of the usefulness of such type of tests.

The results so far obtained from the pilot study version of the tests are reasonably encouraging and it proves that such type of tests can be prepared and can be of use in those situations where verbal intelligence tests fail, due to language handicap on part of the student.

So our next step would be to construct new items similar to the one's found successful in the pilot study and the experimental form of the non-language test would be formed by using these new items along with the 22 old items selected earlier.

This time it is planned to divide the test into five parts as follows : (A) Analogies, (B) Classification, (C) Opposites, (D) Picture Arrangement, (E) Picture Completion. Each part would include near about 50 items and considering the length of these parts they could be administered as separate tests if required.

It is further planned that along with this test the Abstract Reasoning and Verbal Reasoning tests would be administered as done earlier.

It had been pointed out in the very beginning that not necessarily all forms of abstract reasoning items are unrelated to verbal reasoning. More-over though the simple ones may be solved at the perceptual level, yet the more complex ones might need verbalization for their solution. The obtained results also suggest that this assumption might be true as, for some of the items of the Non-Language test, all the three sets of discrimination values are quite high. It seems that *for the next stage* of this study (*i.e.*, during the experimental version) it would be better to conduct an item analysis of the Abstract Reasoning test on the basis of the total score on the Verbal Reasoning Test and a scoring key could be so developed for the Abstract Reasoning Test with the items showing good discrimination values so that these would measure the verbal reasoning through item types generally used in Abstract Reasoning test. After this, two item analyses should be done for the Non-Language test, one on the basis of the total score on Verbal Reasoning and another on the basis of the total score on Abstract Reasoning obtained by using the revised scoring key. The Non-Language Test could then be revised by selecting those items which would show good discrimination values for both the item analyses.

The results obtained with this experimental form would be reported later.

REFERENCES

1. DAVIS, F. B. (1949). Item Analysis Data, their computation, interpretation, and use in test construction, Massachusetts, Graduate School of Education, Harvard University.

2. DOPPELT, JEROME, E. (1954) "Progress in the measurement of Mental Abilities"—Educ. Psychol. Meas., 14, 261–64.

3. FREEMAN, F. N. (1939) *Mental Tests*, London, George, G., Harper & Co., Ltd., 356.

4. HARPER, A. E. Item Analysis Chart—Psychometry Unit, Indian Statistical Institute, Calcutta.

5. HARPER, A. E. Personal Communications.