Indian Journal of Psychology
Vol. 46 No. 3 September, 1971. 235—246

construction and development of a non-language test of verbal intelligence

S. CHATTERJI & MANJULA MUKERJEE

INTRODUCTION

A large number of non-verbal tests of intelligence have made their appearance during the last few years in India and abroad. Research evidence however indicates that it is advisable to use both a verbal and a non-verbal test on which comparable norms are available as an aid to guidance But at the same time it is important to note that the index of ability as measured through performance test and other forms of non-verbal tests is not as stable or reliable as is required in a guidance situation. Moreover, the usual type of non-verbal tests is less efficient in predicting the scholastic success than the available verbal test.

These points were discussed at the "All India Workshop for long Range Planning in the Testing field" held at Bhopal in 1962 and it was suggested by Harper that it was necessary to develop a test which would measure verbal ability through some non-language fair medium. The idea of using a non-language medium was first introduced in the testing field in India by one of the authors in "Chatterji's Non-Language Preference Record Form—962". The items in a non-language test had to be presented in such a way that while mentally solving these items the candidate has to take the help of language. Language would serve only as an intervening variable between the visual presentation of the material and

S. Chatterjee and Manjula Mukherjee are working in Indian Statistical Institute. 203, Barrackpore Trunk Road, Calcutta-35, W. Bengal.

final answering by putting cross marks on the answer sheet. The basic difference between a non-verbal test and this new idea of non-language test lies in the fact that many types of non-verbal items like progressive matrices etc., may be solved at the perceptual level but as defined earlier the later type essentially involves the use of language as intervening factor.

It may be true that non-verbal tests as they in many cases do not involve the use of language, fail to provide adequately accurate prediction of future academic performance when compared to similar verbal tests. Hence, it was reasonably expected that a non-language test would meet the dual challenge of language and accurate prediction and this is of special importance in a multilingual country like India or in a country where emigration poses linguistic problems.

As such a non-language test, if developed could serve several important functions, e. g., could serve as an equating test when different regional versions of a particular test are required to be developed in different languages for nation wise use or could be used to spot out cases where scholastic backwardness is due to linguistic difficulty and not due to lack of ability etc., it was decided to construct and standardize such a test.

To find out whether it was really feasible to develop such a test for measuring intelligence as had been done in the case of interest measurement, a pilot study was undertaken and as encouraging results were obtained (Chatterjee and Mukherjee) an experimental form of the test was prepared and tried out.

2.0 Description of the experimental form of the Non-Language Test (NL)

Before discussing further a description of the item types included in the experimental form and other details are presented below.

2.1. Item types used in the experimental version

(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 have 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) **Opposites**—Here five pictures are presented. The candidate is required to find from among the last four pictures, that one which respresents something which is different from the one shown in the first picture.

(D) **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. (1)

2.2 Other details about the experimental form of NL.

New items of the type found successful earlier, were used along with the successful ones from the pilot study version, in this experimental version of the test. This experimental version consisted of four parts as shown in Table 1.

TABLE 1

Showing total number of items and time allowed in each of the four parts of the experimental form of the NL test.

	Part	No. of items	Time
I	Classification	49	20 minutes
II	Opposites	42	20 minutes
III	Analogy	48	20 minutes
IV	Picture Arrangement	27	30 minutes

Though each part was separately timed yet the candidates were permitted to come back to any of the previous parts.

3.0. Plan of the study

As in the case of the pilot study (1) an Abstract Reasoning (AR) and a Verbal Reasoning (VR) test were administered along with the Non-Language test of Verbal Intelligence. The reason behind using these tests was that the non-language test would be revised on the basis of the relation of its items with the VR and AR scores, and for this reason three separate item analyses of NL test were conducted – viz, on the basis of (i) the total score on the NL test itself, (ii) total score on the VR test,

(iii) total score on the AR test. While revising the NL test, only those items which would show fair discrimination values on the first two item analyses and a poor discrimination value from the third analysis would be retained. This would ensure that the items included in the revised NL test would be internally consistent, the score on the VR and the NL tests would be highly correlated and lastly the NL test would not measure what the AR test was measuring. These two tests viz., Abstract Reasoning & Verbal Reasoning tests were earlier revised on the basis of item analysis against the respective total scores as the criterion. The numbers of items and the time limits for these two tests are presented below.

Abstract Reasoning	33 items	30 minutes
Verbal Reasoning	44 items	30 minutes

4.0 Sample :

Data were to be collected on this form for purposes of item analysis. Hence collection of data on a wide scale was not necessary, but the sample drawn should not be basically different from the group for whom the test was being designed in terms of level etc. So, for this present study five schools were selected from different parts of Culcutta at random and of these five schools, three were for girls and two for boys. The total number of boys and girls in these two groups were 233 and 249 respectively. As done in the pilot study (1) the test was administered on class VIII students of these selected schools. It may be mentioned here that the age range at this level was 11-13 years.

5.0. Analysis of the data

5.1. Mean, Standard deviation, etc.

To start with, the means and standard deviations were calculated on the basis of 100 boys and 100 girls selected at random from the sample obtained. These mean values and standard deviations are presented in Table-2 along with comparable values computed after combining these two groups.

The means and the standard deviations of the total score & the part scores show that the test was quite suited to the level of the group concerned, and it is also evident that sub-test III was the easiest part. Hence while revising the test the serial ordering of the parts has to be changed.

TABLE 2

Showing X and for NL, AR & VR scores obtained on the basis of 100 boys and 100 girls selected at random from the sample obtained.

Groups	Boy's		Girl's		Total		Maximum possible	Critical
Sub-tests	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ	Score	Value
Part-I	35.05	4.14	34.86	4.65	34.96	4.40	49	.03
Part-II	24.77	5.60	25.73	5.55	24.24	5.60	42	1.22
Part-III	34.68	9.23	35.22	7.79	34.95	8.55	48	.45
Part-IV	15.14	5.27	14.91	5.27	15.02	5.27	27	.31
Total	109.64	20.16	110.71	18.82	110.18	19.51	166	.39
AR	18.27	8.31	19.07	7.24	18.67	7.80	33	.73
VR	28.05	6.85	28.00	6.28	28.02	6.57	44	.05

The obtained mean values for these two groups were tested for significance and no significant difference was observed.

5.2. Intercorrelations & reliabilities

The intercorrelations among the subtests, and those among the total NL score, AR & VR were obtained on the basis of the subsample of 200 cases used for calculating the X's and σ's. The obtained values are presented in tables 3 & 4. Reliabilities for the NL subtests, AR & VR were estimated by using the KR-21 formula and the obtained values are also presented in tables 3 and 4.

TABLE 3

Showing intercorrelations among the four NL subtests on the basis of a subsample of 200 cases (100 Boys & 100 Girls)

NL Sub-tests	Part-I	Part-II	Part-III	Part-IV
Part-I	—	.53**	.61**	.51**
Part-II	.53**	—	.57**	.53**
Part-III	.61**	.57**	—	.54**
Part-IV	.51**	.53**	.54**	—
Reliability (KR-21)	.50	.69	.89	.79

* Indicates significant at the 5% level.
** indicates significant at the 1% level.

TABLE 4

Showing intercorrelations among NL total score, AR and VR on the basis of a sub-sample of 200 cases (100 Boys & 100 Girls)

Tests	NL	AR	VR
NL	—	.59**	.65**
AR	.59	—	.61**
VR	.65**	.61**	—
Reliability (KR - 21)	.91	.90	.78

** indicates significant at the 1% level.

The reliabilities of the NL, AR and VR scores were adequately high for all general use. It may also be observed from the figures in table 4 that the correlation between unrevised NL scores and AR and that between AR and VR were high and as discussed earlier it was required to reduce these correlations through successive revisions which would make NL more related to VR than to AR. It may be true that though some of the AR items could be solved at the perceptual level, yet there were some items in solving which language had to be used by the candidates. In such cases it was expected that the correlation between these items and the VR total score would be high. Hence, an item analysis of AR with VR total score as criterion was conducted and those AR items which showed high positive relation with VR score were dropped out while revising AR. With this revised AR scores as the criterion an item analysis of NL was to be done. This revised AR scoring key consisted of 20 items.

5.3. Validity of the Experimental form of NL

Total marks obtained by the students at the class VIII annual examination were used as the criterion and the obtained correlations between this criterion & NL scores are presented in table 5 for each of the school groups & these figures show that even at the unrevised state the NL held great promise.

TABLE 5

Showing the product-moment correlation between test scores and Class VIII annual examination marks separately for each of the five schools.

Schools	NL Part-I	NL Part-II	NL Part-III	NL Part-IV	NL (Total)	AR	VR
1 (N=68)	.17	.11	.17	.03	.04††	.18	.12
2 (N=115)	.05	.18†	.33††	.23††	.28††	.30††	.49††
3 (N=106)	.12	.09	.24†	.20	.22†	.18	.27††
4 (N=113)	.13	.07	.41††	.19†	.28††	.23	.34††
5 (N=73)	.37††	.16	.23†	.01	.25†	.15	.47††
Mav. Possible score	49	42	48	27	166	33	44

†indicates significant at 5% level ††indicates significant at 1% level.

6.0. Item Analysis

It has been mentioned earlier that this Non-Language test was to be of such a type that it would be measuring verbal reasoning i.e., from the practical point of view the correlation between NL scores and VR scores should be of a high order. Moreover it was also felt desirable that the relation between NL and Abstract Reasoning should be as low as possible. Hence to see how far these dual aims were being served by the NL it was planned to conduct item analysis of NL items separately against three criterion scores as follows :

(i) Item analysis of each of the four parts of NL against the total score on that part.

(ii) Item analysis of the NL on the basis of the scores obtained on the revised Abstrat Reasoning test.

(iii) Item analysis of the NL on the basis of the scores obtained on the Verbal Reasoning test.

For item analysis a total of 370 cases (185 boys and 185 girls) were taken at random from the unusable answer sheets of the tested 482 students to facilitate calculation. Upper and Lower 27% of the distribution were used for estimating the discrimination and difficulty values (Harper) Examination of the obtained difficulty indices show that the items in NL were quite suited to the level of the groups concerned as for 31 out of the 166 in the NL test or for only 19% of the total number of items, the difficulty values were below 50. Moreover, it was observed that about 50% of the items in parts I and II had discrimination values above 20, whereas about 95% of the items in parts III and IV had discrimination values above 20 when item analysis was done on the basis of the relevant NL part scores. Hence, it was clear that an internally homogeneous test of adequate length could be formed by revising the experimental version of the test. When item analysis of the NL was done against VR score as the criterion, about 50% of the items in parts III and IV had discrimination values equal to or above 20 and hence these items could be considered as suitable for inclusion in the revised version. For parts I and II the cutting point for screening out unusable items had to be relaxed a bit and by taking a discrimination value of 15 and more as acceptable these two parts could be revised.

While revising the experimental form of NL, first of all, NL items which had a discrimination index of 20 or above with NL score as criterion were picked up. Next, from among these selected items those which had the discrimination index of 10 or above in Parts I and II and 15 or above in Parts III and IV with VR score as criterion were retained. From among those items which survived these two screenings, items whose discrimination indices obtained on the basis of revised AR total score were at least less by 5* than the corresponding one obtained on the basis of total VR scores as criterion were picked up. In this group of items which survived the third screening, it was found that the difference between the indices obtained on the basis of total VR and revised AR scores as criterion, were much greater than 5 in most of the cases. The discrimination values for the selected NL items obtained with AR and VR scores as criterion are presented in table 6.

* It may be mentioned here that this value i.e., 5 was arbitrarily set up and perhaps a higher value of the difference would have been more desirable but as the correlation between the revised AR and VR (as may be seen later) itself was as high as +.52 it would have been difficult to get items which would be highly related to VR and at the same time very poorly related to AR. Hence, this value i.e., 5 was accepted,

TABLE 6

Showing the discrimination values of NL items retained in the revised version, with AR and VR scores as criterion

PART-I			PART-II			PART-III			PART-IV		
Item No.	On the basis of AR	on the basis of VR	Item No.	On the basis of AR	On the basis of AR	Item No.	On the basis of AR	On the basis of VR	Item No.	On the basis of AR	On the basis of VR
4	5	11	3	11	20	2	14	19	3	11	20
9	9	26	4	10	15	5	20	25	4	15	20
11	22	30	6	7	20	8	15	20	7	8	19
16	18	29	7	7	18	16	15	22	10	5	11
21	9	17	9	11	24	23	18	24	12	15	20
24	7	17	13	9	14	25	12	20	15	16	24
27	15	25	14	12	20	26	7	12	16	8	19
29	6	15	17	7	19	27	10	16	18	18	25
32	2	13	24	20	26	30	20	25	19	16	22
34	10	32	26	9	24	31	12	22	20	6	16
36	12	19	29	20	25	33	20	25	23	5	10
38	9	17	31	0	14	34	12	18	24	24	31
44	9	14	38	12	17	36	20	27	25	13	20
45	3	14	42	8	13	39	12	18	26	26	36
						41	10	18			
						42	22	27			
						43	22	27			
						44	7	28			
						46	11	16			
						47	19	27			

7.0. The available information on the revised version of the NL

By utilizing these items, the revised form of the NL test was assembled and the available information on this version on the basis of data on the experimental form is given below.

The mean, standard deviation, & intercorrelations for each part were calculated on the basis of the new set of scores obtained by using the new scoring key of the test. These values are presented in table 7.

TABLE 7

Showing the inter-correlations among different parts of the test with corresponding means, and standard deviations

PART	I	II	III	IV	Total	(Revised) AR	VR
I	—	.38	.50	.39	.70	.31	.46
II	.38	—	.45	.45	.75	.34	.38
III	.50	.45	—	.44	.81	.44	.64
IV	.39	.45	.44	—	.78	.36	.45
Total	.70	.75	.81	.78	—	.48	.63
AR (Rev)	.31	.34	.44	.36	.48	—	.52
VR	.46	.38	.64	.45	.63	.52	—
Mean	9.85	9.11	14.72	7.25	40.93	10.65	28.03
	2.05	2.75	3.45	3.16	8.77	4.44	6.44
Max. possible	14	14	20	14	62	20	45

It may be noted that all these correlations were significantly different from zero.

From Table 7 it can be seen that the inter-correlations among different parts of the NL test are high but on the average these values are lower than the corresponding values of the experimental form. The correlation between AR and VR has decreased from what it had been earlier. The correlation between AR, and different parts of the NL and the total score of NL are not very high i.e., all of them are lower than that between AR and VR (.52), whereas with VR all the parts have sufficiently high correlations.

By the time the NL test was revised the class IX annual examination marks for the students who took the experimental from of the NL were available and the product-moment correlation of the NL test scores calculated against this criterion is presented in table 8.

TABLE—8

Showing the obtained validity coefficients for the NL parts scores, total score, AR and VR for different groups included in the sample.

Tests / Groups	I	II	III	IV	Total	AR	VR
	r_1	r_2	r_3	r_4	r_5	r_6	r_7
Boy's Science (N = 60)	.40**	.39**	.60**	.29**	.60**	.67*	.70*
Boy's Humanities (N = 17)	.38	.45	—.03	—.08	.12	—.02	—.58**
Boy's Technical (N = 25)	.31	.34	.10	.14	.30	.35	.38**
Boy's Commerce (N = 61)	.15	.09	—.06	.19	.10	—.06	.07
Girl's Science (N = 90)	.32**	.34	.43**	.21*	.45**	.39**	.45**
Girl's Humanities (N = 83)	.50**	.22*	.30**	.26**	.53**	.27*	.70**
Girl's Home Science (N = 11)	.14	.19	.62*	.14	.44	.25	.07
Girl's Commerce (N = 10)	.11	.02	.09	.00	.00	.02	.02
Max. Possible Score	14	14	20	14	62	20	45

**indicates significant at the 1% level.

*indicates significant at the 5% level.

The values presented in table 8 show that the NL hold great predictive potentiality specially for both the science groups and also for the Girls' Humanities group. For the Boy's Technical and Girls' Home Science groups nothing definitely could be said as the number of cases involved is too small, yet the trends observed in these cases are highly encouraging. The obtained values for the Boy's Humanities group are rather discouraging.

8.0. Item Analysis of the revised NL test :

The answer sheets of the experimental form of the NL were rescored by using only those items which were included in the revised version. By using these new scores as the criterion an item analysis of NL was conducted and the obtained values are presented in Table 9.

TABLE 9

Showing the difficulty and discrimintion values of the items included in different parts of the revised version of the test on the basis of the respectivc part scores.

Part-I (Now part-II)			Part-II (Now Part-III)			Part-III (Now Part-I			Part-IV (Now Part-IV)		
Item No.*	Diff	Disc	Item No.*	Diff	Disc.	Item No.*	Diff	Disc.	Item No.*	Diff	Disc.
4	90	20	3	55	49	2	81	40	3	55	45
9	67	34	4	73	45	5	57	43	4	88	24
11	50	48	6	65	55	8	52	34	7	53	40
16	58	64	7	58	35	16	64	42	10	38	27
21	64	45	9	59	55	23	48	30	12	38	55
24	68	37	13	36	37	25	69	38	15	39	60
27	65	37	14	65	29	26	74	30	16	48	46
29	79	20	17	58	29	27	47	30	18	58	40
32	82	20	24	53	44	30	79	40	19	56	44
34	36	21	26	60	50	31	70	30	20	46	46
36	75	45	9	65	42	33	75	30	23	51	37
38	32	37	31	59	45	34	59	35	24	33	47
44	61	44	38	81	40	36	75	40	25	69	42
45	76	27	42	53	28	39	77	45	26	47	53
						41	82	40			
						42	64	45			
						43	38	43			
						44	71	27			
						46	62	27			
						47	61	45			

*The original item number of the experimental version of the test

The values presented in table 9 show that the discrimination values of the revised NL test with NL part, score as criterion are sufficiently high and the test is not difficult for the group for which it is meant. The revised NL test is being printed and on the basis of data collected on the revised version regional norms would be developed and further follow up studies would be conducted. These would be reported in due course.

REFERENCES

CHATTERJI, S. AND MUKHERJEE, M. The Development of Non-Language test of verbal Intelligence. *Journal of Psychological Researches*, 1967, Vol. 11, No. 2, pp. 58-68.

HARPER, A.E. Item Analysis Chart—Psychometry Unit, *Indian Statistical Institute*, Calcutta-35.

HARPER, A.E. *Personal Communications.*