CHAPTER V

*
*
* * * STANDARDIZATION OF THE TEST

The final test, as described in the preceding chapter, has eight items. Now it is essential to see whether the test is valid and reliable, i.e. whether it would actually measure what it is meant to measure, and whether it would give reliable results if given time and again. Further, if so, it would be desirable to fix the procedure of scoring and interpreting, i.e. to establish the norms - for purpose of interpretation of results obtained on the test. All this means that the test should be standardized.

WHAT IS STANDARDIZATION ?

By standardization we mean "uniformity in the procedure of administration and scoring of the test."

The process of standardization of a test involves, also among other things, the standardization of method, material and results. The standardization of methods procedes the formulation of the try-out test or the pilot work. The standardization of material is done practically

through the analysis of items after administering the tryout form of the test. The standardization of results is done after the administration of the final form of the test.

The standardisation of the method and material already done, the investigator had now to do standardization of results only. Thus, now the standardization includes preparation of norms and finding out of validity and reliability.

The first step is to see whether the test can be administered within a reasonable time period without bringing in an element of fatigue. It was easy to determine this because the investigator by now knew the amount of time each item takes. She calculated and found out that it took 5 minutes for 'Uses Test' ($2\frac{1}{2}$ minutes for each item), 20 minutes for 'Creative Writing Test' (10 minutes for each item), 10 minutes for 'Consequences Test' (5 minutes for each item), and 10 minutes for 'Problem Solving Test' (5 minutes for each item). The total time for four sub-tests each having two items thus comes out to be about one hour which includes test administration time too. Let us now turn our attention to the procedure of standardization or establishing norms, validity and reliability.

NORMS

A raw score of an individual has no meaning until it is comparable with other members of the group. This comparison is provided by norms. By norms we mean a standard behaviour of the members of a group in reference to characteristics like age, sex, achievement, etc. Norms represent descriptive frame-work for interpreting the scores of an individual or a group.

To find the norms the test had to be administered to a large sample. This sample was selected from Delhi. A complete list of Delhi's schools was consulted for the selection of the sample.* Delhi school network is divided into six zones. Each zone has boys' as well as girls' schools. It also has Government, Municipal, Private and Public Schools. Crosswise it includes the urban and rural schools.

SAMPLE FOR THE NORMS

A sample of 1000 students was selected for preparation of norms. They were selected from various kinds of schools. It was important to lay down a criteria under which the

---

* The list of distribution of different types of schools in Delhi is given in Appendix I.

students could be selected. As stated earlier, Delhi has various kinds of schools. The table No.7 given below indicates the number and percentage of schools selected by the investigator:

Table 7: Showing Distribution of Sample with Reference to Kinds of Schools

	Government	Private	Public	Others	
Total	249 (52%)	141 (29%)	11 (23%)	67	
Selected	9	5	2	Rural 1	Municipal 1

It may be seen that 18 schools* were taken in all, 9 from government schools, 5 from private schools, 2 from public schools, 1 from municipal schools and 1 from rural school. These included three co-educational schools.

A random selection of subjects was made out of 9th, 10th and 11th class pupils because the age group - 14 to

---

* The list of schools is given in Appendix J.

16 years - required students reading in these classes. The test was administered to the pupils in groups during the year 1971, following all necessary precautions.

SCORING

In the pilot study the test was administered to 350 students and a certain number of responses as stated in chapter III were received.* The complete final form of the test was administered to 1000 students. The responses received indicated that the variety of responses had increased. This necessiated the formation of more categories. This, of course, was done. The responses were scored. After scoring the papers, two master sheets were prepared which contained all the data for boys and girls separately i.e. scores achieved on all the eight items measuring fluency, flexibility, originality, organisation, imagination and richness.

The pattern in which the master sheets for 1000 students were prepared is given below in table No.8 .

---

* In the last chapter we discussed that one of the tests i.e. test no.1 'Sentence Completion Test' which did not satisfy the condition was dropped. Thus, the number of tests was only four. For the sake of clarity the older serial number has been retained. However, this test (no.1) is not mentioned at any place in this Chapter.

The table indicates the details of the master sheet for one student. Similar sheets were prepared for all the 1000 students. The sheets constitute a huge bulk of papers and has not been placed in the text of the thesis as it was not considered necessary to do so.

Table 8: Showing the Pattern of the Scores Written in Master Sheet

Sr. No.	Preliminaries	Item	Fluency	Flexibility	Originality	Organisation	Imagination	Richness
1	Name: Saroj Bala Class: X Age: 15 years Sex: Female School: Govt. Girls' Higher Secondary School Roll No.: 8	I	4	3	-	-	-	-
		II	6	5	-	-	-	-
		III	-	-	1	4	2	3
		IV	-	-	-	2	3	2
		V	16	8	4	-	-	-
		VI	9	3	-	-	-	-
		VII	6	4	4	-	-	-
		VIII	9	8	4	-	-	-
	Total		40	33	13	6	5	5

The total score say for fluency in the case of this girl, is 40. This was obtained by adding up scores on all items for fluency e.g. $4+6+16+9+6+9 = 40$. Similarly,

scores on all items for flexibility, originality, organization, imagination and richness were added. In the case of this girl, the score for fluency, originality, organization, imagination and richness were 40, 33, 13, 6, 5 and 5 respectively as shown above. These scores were called raw scores. These scores were further segregated in respect of sex. e.g. 14 years girls and 14 years boys, 15 years girls and 15 years boys, 16 years girls and 16 years boys. All the data* were thus divided into six groups. These raw scores as they were could not be interpreted for the following reasons :

(1) The number of responses of each student differed from item to item.

(2) The number of responses of different students again varied for each item.

This presented a position under which there was no reference for a comparison. If, for instance, there were a position under which total number of responses would be say, 50, the number of responses of an individuals to an item could be interpreted as it was. Since such a position did not exist, it became necessary to convert this raw score into a standard score and interpret it to get a view

---

* These data being very bulky, have not been included in the text nor in the appendices as it was not considered necessary to do so.

of the creativity of an individual. The raw score was therefore converted into T scores. The following standard method was applied for conversion. To get a T score, T scale was constructed as shown in different columns of tables 9-14 for different scores on creativity. Frequencies of each score in column No.1 were found as in column No.2 and their cummulative frequencies prepared as seen in column No.3 . The following formula was then applied to obtain specific value in columns No.4 & 5.

$$\frac{\text{Cummulative Frequency Below Score} + \frac{1}{2} \text{ on Given Score}}{N} \times 100$$

In other words $\frac{1}{2}$ of frequency of each test score was added to the cummulative frequency below that test score as seen in column no.4 . This score when divided by total number of cases (N) and multiplied by 100 gave percentage of each score as seen in column 5. These percentages were then converted into T-scores with the help of table G given by Garrett[1] in his book 'Statistics in Psychology and Education'. This conversion of all raw scores into T-scores is given in column 6. Raw scores are again given in column 7 along side T-scores in column 6 for user's convenience. All these steps of scores have been presented in different columns, as described above, in tables no. 9, 10, 11, 12, 13

---

1 Garrett, H.E.: Statistics in Psychology and Education, Longman, Green & Co., London, 1960, p.455.

and 14 respectively for raw scores on different aspects of creativity, viz. fluency, flexibility, originality, organization, imagination and richness.

Table No.9: Showing Raw Scores and T-Scores on Fluency

Test Score	f	Cum f	Cum f below given score + ½ on given score	%	T-score	Test score
1	2	3	4	5	6	7
57	1	700	699.5	99.93	82	57
56	1	699	698.5	99.79	79	56
55	0	698	698.0	99.71	78	55
54	0	698	698.0	99.71	78	54
53	0	698	698.0	99.71	78	53
52	1	698	697.5	99.64	77	52
51	0	697	697.0	99.57	76	51
50	1	697	696.5	99.50	76	50
49	1	696	696.5	99.36	75	49
48	0	695	695.0	99.29	75	48
47	0	695	695.0	99.29	75	47
46	3	695	693.5	99.07	74	46
45	1	692	691.5	98.78	73	45
44	1	691	690.5	98.64	72	44

Continued..

Test Score	f	Cum f	Cum f below given score + ½ on given score	%	T-score	Test score
1	2	3	4	5	6	7
43	3	690	688.5	98.36	71	43
42	4	687	685.0	97.86	70	42
41	5	683	680.5	97.21	69	41
40	8	678	674.0	96.29	68	40
39	9	670	665.5	95.07	68	39
38	8	661	657.0	93.86	65	38
37	12	653	647.0	92.43	64	37
36	15	641	633.5	90.50	63	36
35	16	626	618.0	88.29	62	35
34	15	610	602.5	86.07	61	34
33	23	595	583.5	83.36	60	33
32	31	572	556.5	79.50	58	32
31	19	541	531.5	75.93	57	31
30	26	522	509.0	72.71	56	30
29	20	496	486.0	69.43	55	29
28	41	476	455.5	65.07	54	28
27	41	435	414.5	59.21	53	27
26	48	394	370.0	52.86	51	26
25	47	346	322.5	46.07	49	25

Continued...

Test Score	f	Cum f	Cum f below given score + ½ on given score	%	T-score	Test score
1	2	3	4	5	6	7
24	40	299	279.0	39.86	47	24
23	27	259	245.5	35.07	46	23
22	36	232	214.0	30.57	45	22
21	33	196	179.5	25.64	43	21
20	33	163	146.5	20.93	42	20
19	17	130	121.5	17.36	41	19
18	23	113	101.5	14.50	39	18
17	16	90	82.0	11.71	38	17
16	15	74	66.5	9.50	37	16
15	11	59	53.5	7.64	36	15
14	12	48	42.0	6.0	35	14
13	9	36	31.5	4.50	33	13
12	8	27	23.0	3.29	32	12
11	7	19	15.5	2.21	30	11
10	5	12	9.5	1.36	28	10
9	4	7	5.0	0.71	25	9
8	2	3	2.0	0.29	22	8
7	1	1	0.5	0.07	18	7

Table 10: Showing Raw Scores into T-Scores on Flexibility

Test Score	f	Cum f	Cum f below given score + ½ on given score	%	T-score	Test Score
1	2	3	4	5	6	7
46	1	700	699.5	99.93	82	46
45	1	699	698.5	99.79	79	45
44	0	698	698.0	99.71	78	44
43	0	698	698.0	99.71	78	43
42	0	698	698.0	99.71	78	42
41	0	698	698.0	99.71	78	41
40	0	698	698.0	99.71	78	40
39	0	698	698.0	99.71	78	39
38	1	698	697.5	99.64	77	38
37	0	697	697.0	99.57	76	37
36	2	697	697.0	99.43	76	36
35	2	695	694.0	99.14	74	35
34	4	693	691.0	98.71	72	34
33	4	689	687.0	98.14	71	33
32	1	685	684.5	97.79	70	32
31	7	684	680.5	97.21	69	31
30	10	677	672.0	96.00	68	30
29	12	667	661.0	94.43	66	29

Continued....

Test Score	f	Cum f	Cum f below given score + ½ on given score	%	T-score	Test Score
1	2	3	4	5	6	7
28	14	655	648.0	92.57	64	28
27	14	641	634.0	90.57	63	27
26	29	627	612.5	87.50	62	26
25	22	598	587.0	83.86	60	25
24	33	576	559.5	79.93	59	24
23	34	543	526.0	75.14	57	23
22	42	509	488.0	69.71	55	22
21	41	467	446.5	63.79	54	21
20	63	426	394.5	56.36	52	20
19	38	363	344.0	49.14	50	19
18	36	325	307.0	43.86	48	18
17	54	289	262.0	37.43	47	17
16	42	235	214.0	30.57	45	16
15	30	193	178.0	25.43	43	15
14	28	163	149.0	21.29	42	14
13	36	135	117.0	16.71	40	13
12	28	99	85.0	12.14	38	12
11	24	71	59.0	8.43	36	11
10	16	47	39.0	5.57	34	10

Continued...

Test Score	f	Cum f	Cum f below given score + ½ on given score	%	T-score	Test Score
1	2	3	4	5	6	7
9	15	31	23.5	3.36	32	9
8	11	16	10.5	1.50	28	8
7	4	5	3.0	0.42	24	7
6	1	1	0.5	0.07	18	6

Table 11: Showing Raw Scores into T-scores on Originality

Test Score	f	Cum f	Cum f below score + ½ on given score	%	T-score	Test Score
1	2	3	4	5	6	7
32	2	700	699.0	99.86	80	32
31	1	698	697.5	99.64	77	31
30	1	697	696.5	99.50	76	30
29	0	696	696.0	99.43	75	29
28	1	696	695.5	99.36	75	28
27	0	695	695.0	99.29	75	27
26	1	695	694.5	99.21	74	26
25	0	694	694.0	99.14	74	25
24	1	694	693.5	99.07	74	24
23	2	693	692.0	98.86	73	23

Test Score	f	Cum f	Cum f below score + ½ on given score	%	T-score	Test Score
22	3	691	689.5	98.50	72	22
21	2	688	687.0	98.14	71	21
20	2	686	685.0	97.86	70	20
19	0	684	684.0	97.71	70	19
18	1	684	683.5	97.64	70	18
17	4	683	681.0	97.29	69	17
16	5	679	676.5	96.64	68	16
15	5	674	671.5	95.93	67	15
14	6	669	666.0	95.14	67	14
13	10	663	658.0	94.00	66	13
12	6	653	650.0	92.88	65	12
11	12	647	641.0	91.57	64	11
10	20	635	625	89.28	62	10
9	17	615	606.5	86.64	61	9
8	32	598	582.0	83.14	60	8
7	38	566	547.0	78.14	58	7
6	47	528	504.5	72.07	56	6
5	63	481	449.5	64.21	54	5
4	66	418	385.0	55.00	51	4
3	62	352	321.0	45.86	49	3
2	94	290	243.0	34.71	46	2
1	67	196	162.5	23.21	43	1
0	129	129	64.5	9.21	37	0

Table 12: Showing Raw Scores into T-scores on Organization

Test Score	f	Cum f below score + ½ given score		%	T-score	Test Score
10	2	700	699.0	99.85	80	10
9	2	698	697.0	99.57	76	9
8	6	696	693.0	98.99	73	8
7	17	690	681.5	97.35	69	7
6	24	673	661.0	94.42	66	6
5	47	649	625.5	89.36	62	5
4	78	602	563.0	80.42	59	4
3	91	524	478.5	68.35	55	3
2	109	433	378.5	54.07	51	2
1	108	324	270.0	38.57	47	1
0	216	216	108.0	15.42	40	0

Table 13: Showing Raw Scores and T-scores on Imagination

Test Score	f	Cum f below score	Cum f below + ½ given score	%	T Score	Test Score
8	2	700	699.0	99.85	80	8
7	6	698	695.0	99.28	75	7
6	8	692	688.0	98.29	71	6
5	13	684	677.5	96.78	68	5
4	25	671	658.5	94.07	65	4
3	78	646	607.0	86.71	61	3
2	145	568	495.5	70.78	55	2
1	136	423	355.0	50.71	50	1
0	287	287	143.5	20.50	42	0

Table 14: Showing Raw Scores and T-scores on Richness

Test Score	f	cum f	Cum f below given score + ½ on given score	%	T-score	Test Score
9	1	700	699.5	99.93	82	9
8	1	699	698.5	99.78	78	8
7	2	698	697.0	99.57	76	7
6	15	696	688.5	98.36	71	6
5	23	681	669.5	95.64	67	5

Continued.....

Test	f	Cum f	Cum f below given score + ½ on given score	%	T Score	Test Score
4	95	658	610.5	87.21	61	4
3	113	563	506.5	72.36	56	3
2	225	450	337.5	48.21	49	2
1	114	225	168.0	24.00	43	1
0	111	111	55.5	7.93	36	6

The T-scores of a 15 years old girl is given below as an example.

Table 15: Showing for Reference the T-score of a 15 Year Old Girl

	Fluency	Flexibility	Originality	Organization	Imagination	Richness	
Raw Score	40	33	13	6	5	5	
T-Score	68	71	66	66	68	67	= 406

In the case of this girl the total T-score is 406.

Thus, the total score of each student of the sample consisting of 1000 students were computed. Having constructed the T-scores and the total scores it was now necessary to find mean scores of all the six groups i.e. 14 year boys, 14 year girls, 15 year boys, 15 year girls, 16 year boys and 16 year

girls. The following formula was used for computing the mean scores.

$$\text{Mean} = AM + \frac{\Sigma f x'}{N} \times i \ldots\ldots$$

where

AM = Assumed Mean

Σ = Sum

f = Frequency

x' = Deviation from Assumed Mean

N = Total no. of cases

i = Class-interval

To find out dispersion in each mean, standard deviations were computed. The following formula was used to compute standard deviation.

$$SD = i\sqrt{\frac{\Sigma f x'^2}{N} - \left(\frac{\Sigma f x'}{N}\right)^2}$$

where symbols are the same as given above. The mean scores and SD of each of six groups are summarized in table no.16 below.

Table 16: Showing M and SD for each Group

Group	14 year boys	14 year girls	15 year boys	15 year girls	16 year boys	16 year girls
Mean	305.0	299.3	305.2	300.84	297.5	291.3
SD	39.8	45.1	42.3	47.6	41.7	42.2

It would be seen from table 16 that while mean scores for boys were higher than those for girls at each age, the dispersion for girls was higher than that for boys.

Further, the mean scores thus computed showed that on the whole 15 year olds scored the highest and 16 year olds the lowest, and on the whole boys scored higher than girls. Groupwise, 15 year old girls stood highest almost with 14 year old boys, while 16 year old girls stood lowest and all others almost equal in between.

It is very obvious that the mean score of boys is more than that of girls, while the SD is more in the case of girls, meaning that boys were higher on creativity and less variable or more homogeneous in distribution than girls.

Finding out mean and standard deviation was necessary but not enough. It was important to find out whether the difference between means was significant in three age groups and two sex groups. To examine the significance of difference between any two means at a time, 't' ratios were computed and it was found from results in tables 17-20 below that boys scored significantly higher than girls, and that 16 year olds were significantly different from 14 and 15 year olds who were mutually not significant.

A STUDY OF AGE AND SEX DIFFERENCES

In order to study whether age and sex made any differences in scores on creativity, a statistical technique called 't' test was applied to the data to examine the significance of differences between any two means at a time. The following formula was used to compute 't' ratio .

$$t = \frac{M_1 - M_2}{\sqrt{\left(\frac{\Sigma x_1^2 + \Sigma x_2^2}{N_1 + N_2 - 2}\right)\left(\frac{N_1 + N_2}{N_1 N_2}\right)}}$$

where

t = t-ratio

M_1 = Mean of the First Sample

M_2 = Mean of the Second Sample

Σx_1^2 = Sum of Squares of Deviation from Actual Mean of First Sample

Σx_2^2 = Sum of Squares of Deviation from Actual Mean of Second Sample

N_1 = Total Number of Cases in First Sample

N_2 = Total Number of Cases in Second Sample

The results obtained have been summarized in the following tables no.17-20.

Table 17: Showing 't' Ratio Between 14 years and 15 years

Age Groups	't' Ratios	Significance
14 years total and 15 year total	.74	NS
14 year male and 15 year males	.95	NS
14 year females and 15 year females	1.51	NS
14 year males and 15 year females	.05	NS
14 year females and 15 year males	.43	NS

All these ratios indicate that the difference between the two age groups anywhere is not significant.

Table 18: Showing 't' Ratios Between 14 Years and 16 Years

Age Groups	't' Ratios	Significance
14 year total and 16 year total	3.71	Sig. .01
14 year males and 16 year males	2.58	Sig. .01
14 year females and 16 year females	1.96	Sig. .05
14 year males and 16 year females	3.14	Sig. .01
14 year females and 16 year males	2.56	Sig. .05

The ~~three~~ ratios above indicate that the differences between the two groups are significant at .01 level and other two differences significant at .05 level.

Table 19: Showing 't' Ratios Between 15 Years and 16 Years

Age Groups	't' Ratios	Significance
15 year total and 16 year total	6.81	Sig. .01
15 year males and 16 year males	2.08	Sig. .05
15 year females and 16 year females	2.02	Sig. .05
15 year males and 16 year females	2.67	Sig. .01
15 year females and 16 year males	1.80	NS

The two ratios indicate that the differences between the two age groups were significant at .01 level while the two ratios were significant at .05 level. Only non-significant is the difference between 15 year females and 16 year males.

Further, 't' ratios were available in above tables to study sex differences at different age levels, but not at same age levels.

Hence, 't' ratios for total groups of boys and girls as at three age levels were computed. All these are separately reproduced in table no.20 .

Table 20: Showing Sex Differences

Age Groups	't' Ratios	Significance
All males and all females	3.07	Sig. .01
14 year males and 14 year females	1.98	Sig. .05
15 year males and 15 year females	3.32	Sig. .01
16 year males and 16 year females	2.53	Sig. .05

All the four ratios are significant. 't' ratios between total males and females as well as 15 year males and 15 year females were significant at .01 level, while the other two were significant at .05 level.

These data indicate that males and females should be treated as two groups for all three age groups viz. 14 years, 15 years, and 16 years.

Looking to the above results, it seemed advisable to prepare separate norms as under, viz. (i) 14 years and 15 years as one group and (ii) 16 years another group. And in each, there should be separate norms for boys and girls. In short, norms were prepared for ready reference in case of the following groups:

(1) Norms for 14 years & 15 years boys.

(2) Norms for 14 years & 15 years girls.

(3) Norms for 16 years boys

(4) Norms for 16 years girls

The author has derived from the raw scores two types of norms for comparison of performances, viz. (i) percentile or decile norms and (ii) standard norms, i.e. Z-scores and T-scores; separately in case of above four groups of sex and age. All these have been presented in tables no.21.

PERCENTILE OR DECILE NORMS

Since, there were no significant differences between 14 years and 15 years on the whole or even sexwise, the investigator decided to prepare norms for 14 years and 15 years as one group and 16 years as a separate group. She constructed percentile norms, using percentile (or decile) rank method for preparation of norms. This method is relatively easy to interpret and most commonly used in creativity tests.

According to Thorndike and Hagen, "Percentile norms are very adaptable and applicable. They can be used wherever an appropriate normative group can bebbtained to serve as a yardstick. They are appropriate for young and old, educational or industrial situation." [1]

1 Thorndike,R.L. & Hagen,E.P.: Measurement and Education in Psychology and Education, John Willey Sons Inc., New York, 1955.

"Percentile method is a graphic way of fixing the point of reference. An individual's percentile rank on a test designates the percentage of cases or scores lying below it. By this, we mean, a person's relative status or position in the hierarchy can be established with respect to the traits or functions being tested. The percentile rank designates one-hundredth part of a distribution, while decile rank designates the 1/10 th part in which any tested person is placed by his score. The term decile technically means a dividing point. Decile rank signifies a range of scores between two dividing points." [1]

A table giving the decile ranks would help in placing a person in his group. His placement would indicate the trend of his creativity. The decile ranks were therefore prepared. The formula[2] used to compute percentile or decile rank is given as follows:

$$P = L + \frac{P_N - f}{f_p} \times i$$

where

P = Percentage of the Distribution Wanted.

L = Lower Limit of the Class-interval.

P_N = Part of N to be counted off in order to reach P.

f = Number of Scores within the Interval upon which P falls.

i = Length of the Class-interval.

---

1 Garrett,H.E.: Statistics in Psychology and Education, Longman Green & Co., 1954, P.64.

2 Garrett,H.E.: Statistics in Psychology and Education, Longman Green & Co., 1964, P.65.

In view of the considerations noted earlier regarding significance of difference between various group, the decile ranks were computed separately for the following four groups for reference.

(1) Combined group of 14 and 15 year males.

(2) Combined group of 14 and 15 year females.

(3) Males of 16 years.

(4) Females of 16 years.

These are represented in table 21.

The raw score of any individual from any of the original six groups can be referred to the respective tables for finding the equivalent percentile score for comparison.

STANDARD NORMS OR T-SCORES

For comparison purpose, still more reliable norms are standard scores. Since the units of a score system based on percentile are so clearly not equal, we look for some other unit that does have the same meaning throughout the whole range of values. Scores in terms of standard deviation units have been developed to serve this purpose. These are Z-scores, T-scores or stanine scores. A Z-score

Table 21: Giving all the Decile Ranks in Detail

Sr. No.	Group	P 10	P 20	P 30	P 40	P 50	P 60	P 70	P 80	P 90	P 100
1	Combined Group of 14 and 15 yr. Males	250.8	266.5	291.8	294.0	302.7	315.8	326.8	340.0	356.6	454.5
2	Combined Group of 14 and 15 yr. Females	240.7	260.7	271.2	282.3	308.1	313.5	326.6	341.3	362.8	404.5
3	Males of 16 years	240.9	261.9	274.7	285.3	292.6	302.3	316.2	336.1	358.3	414.5
4	Females of 16 yrs.	237.0	251.1	262.3	276.5	290.7	300.0	312.0	328.5	349.5	404.5

means so many units of standard deviation below or above the mean. This value may thus be negative or positive and may be in decimal fraction also. Hence to do with the negative sign and fractional value, standard scores are developed with respect to some convenient mean and standard deviation. If the group scores are not normally distributed as assumed, the standard scores are normalized and used for comparison in all cases. Such normalized standard scores with 50 as mean and 10 as standard deviation are specifically called T-scores. The common formula used is:

$$T\text{-score} = 50 + 10 (X - M) / S.D.$$

The other way to derive T-score from frequency has been described earlier. Further, a stanine scale is a condensed T-scale, running from 1 to 9 with 5 as median and .5 as S.D. Any of these standard scores can be used for comparison. The investigator has, in addition to decile ranks, derived the T-scores for raw scores in each aspect of creativity for the same four groups as mentioned earlier. These are presented in table 22 ~~below.~~

RELIABILITY

Reliability refers to the extent to which an instrument yields consistent results on testing and retesting i.e how dependable is it for predictive purposes. If a test does not have a high degree of reliability, it can have but limited value, if any, in predicting an individual's future performance or level of development. Reliability is not, however, an all-or-none law; it is a matter of degree. It is difficult to obtain 100% reliability as far as human behaviour is concerned. There is always some variation in the results, either in terms of internal consistency or of prediction. It may be due to some error of measurement large or small, or to the fact that it is 'normal' for human beings who are constantly in the process of growth and development to vary in performance.

The following methods are normally used to obtain reliability.

(1) The same form of the test may be administered twice to the same group of individuals (test-retest method).

(2) Two separate but equivalent forms of the test may be administered to the same individuals (equivalent form method).

(3) The test items of a single test are subdivided into two presumably equivalent and separate sets (split half method).

Out of the three, the most useful method for a heterogenous test is of the retest variety. The test is repeated on the same group, after a period of time and coefficient of correlation is calculated between the two sets of scores. The advantage of retest is that it yields information about the stability of rank orders of individual's over a period of time. A high correlation from this source indicates that persons change very little in status within their population from the first to the second testing. It also indicates that the test measures the same functions before and after the interval. [1]

The present test was readministered after one week to find the consistency of the test on two occasions. One hundred students were selected for the test and retest. Their scores were tabulated for all the six aspects and converted into T-scores as indicated in chapter IV.

---

1 Gulford,J.P.: Fundamental Statistics in Psychology and Education, McGraw Hill Book Co., P.446-447.

Coefficient of correlation between the two scores was found out with the help of Person's Method of Product-Moment Correlation Coefficient. The formula applied was:

$$r = \frac{\Sigma XY - C_x C_y}{N \sigma_x \sigma_y}$$

where

r = Correlation Coefficient.

X = Scores on First Testing.

Y = Scores on Retesting.

N = Total Number of Cases.

Σ = Sum.

$$C_x = \frac{\Sigma fx'}{N} = \frac{-40}{100} = -.40$$

$$C_y = \frac{\Sigma fy'}{N} = \frac{-29}{100} = -.29$$

$$\sigma_x = \sqrt{\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2} = \sqrt{\frac{1060}{100} - \frac{-40}{100}^2} = 3.23$$

$$\sigma_y = \sqrt{\frac{\Sigma fy'^2}{N} - \left(\frac{\Sigma fy'}{N}\right)^2} = \sqrt{\frac{989}{100} - \frac{-29}{100}^2} = 3.14$$

$$r = \frac{\frac{781}{100} - (-.40 \times -.29)}{100 \times 3.23 \times 3.14} = .75$$

The reliability index as worked out above is .75 . This is a fairly high coefficient of correlation, and shows that the test is reliable.

VALIDITY

An index of validity shows the degree to which a test measures what it purports to measure. The construction and use of a test implies that the instrument has been evaluated against accepted standard, or other criteria which are regarded by experts as the best evidence of the traits or abilities to be measured by the test.[1]

But the problem of validating the test as constructed by the investigator was difficult, because there was no suitable single criteria against which the test could be validated. In this connection it may be stated that since work on creativity is yet in its infancy, no such test or criteria has been developed so far. The investigator, therefore, decided to use a variety of established criteria to validate the test. These were: (a) Torrance Test of Creative Thinking, (b) Shanker's on-the-spot Painting and Writing Competition, (c) Teacher's Rating, (d) Intelligence Test (to make sure that the test measures creativity and not intelligence).

---

1 Freeman, Frank's: Theory and Practice of Psychological Testing, Holt and Rinehart and Winston Inc., New York, 1960, P.26.

TORRANCE TEST OF CREATIVITY THINKING

It is necessary to say a few words about each of these criteria. Torrance Test of Creative Thinking[1] is a well-known test and need not be over-emphasized. It constitutes of verbal and figural tests. Each test has two forms: A & B. The investigator used the verbal test because the test she has prepared is also verbal. Of these two forms A & B, the investigator took up form A. This form could not be used as such for Indian students. It was therefore necessary to translate the instructions and allow the students to write in Hindi. Dr. Torrance himself allows the translation of instructions into Indian languages on the ground that it does not make much difference. But originality score, he says, should be found out for each culture. Accordingly the category system as developed by Dr. Torrance was used while originality score was re-established.[2]

Both the tests i.e. Torrance Test of Creative Thinking and the test constructed by the present

---

1 Torrance Test is given in the Appendix K.

2 Torrance, E.P.
Technical Norms Manual for the Torrance Test of Creative Thinking (Research Edi.).

investigator were administered to a sample of 100 students representing boys as well as girls. Correlations were computed between total scores on the two tests. Correlations were also computed between scores on different aspects measured by the two tests. Person's product moment correlation coefficient formula was utilized, viz.

$$r = \frac{\Sigma XY - C_X C_Y}{N \sigma_x \sigma_y}$$ as noted earlier.

where X and Y would represent scores on the present test, and Torrance Test respectively in this case. A scattered diagram was prepared to find correlation between the two totals. Following is the result :

$C_X = -1.25$ $\quad \Sigma x'^2 = 1.56$

$C_Y = 1.07$ $\quad \Sigma y'^2 = 1.14$

$\sigma_x = 2.67$ $\quad \Sigma XY = 635$

$\sigma_y = 2.11$

$$r = \frac{\frac{635}{100} - (-1.75 \times 1.07)}{100 \times 2.67 \times 2.11} = .72$$

where

X = Stands for investigator's test scores.

C_X = Stands for correction on investigator's test score.

σ_x = Stands for standard deviation of investigator's scores.

Y = Stands for scores on Torrances test of creative thinking.

C_y = Stands for correction on Torrance test of creative thinking score.

σ_y = Stands for standard deviation of Torrance test scores.

Σxy = Sum of products of scores on the two tests.

N = Stands for total number of subjects in the sample.

In addition, correlation between

Fluency of the two tests was found to be = .87

Flexibility of the two tests was found to be = .69

Originality of the two tests was found to be = .75

It is very obvious that fluency, flexibility and originality measured by the two tests w are highly related. The other three aspects i.e. imagination, organization and richness, having amongst themselves comparatively higher correlations, have comparatively low correlations with fluency but not so low with flexibility or originality. These are given below :

Correlation between

Fluency and organization = .33

Fluency and imagination = .15

Fluency and richness	= .49
Flexibility and organization	= .43
Flexibility and imagination	= .28
Flexibility and richness	= .40
Originality and organization	= .34
Originality and imagination	= .44
Originality and richness	= .41

(i) SHANKAR'S PAINTING AND WRITING COMPETITION

Shankar is one of the most celebrated artists and journalists of India. Over twenty years ago he organized a national competition of child artists and a little later of child creative writers. Over a decade now he has been organizing the competition at national as well as international levels. Several thousand students from all over the world participated in the competitions.

Two kinds of competitions were held. One gave option to a child to prepare a painting at home and the other asked him to prepare it on-the-spot. The competition was open to children upto the age of 16 years.

Great accredited writers and artists were actively associated with the examination and judgment of writings

and pieces of art. The award winners from any country in the world are given prizes. These constitute the President of India Prize, the Prime Minister of India Prize, and so on. The competition is held in high esteem both in India and abroad. The investigator, therefore, considered this as a reliable reference against which her test could be validated. This has been a unique approach to validity. The investigator, therefore, got in touch with Mr. Shankar and identified some students in the age group 14 years to 16 years who were awarded the prizes in the year 1970-71. Through his kind assistance twelve students could be traced through their schools and the test was administered to them. The rest had gone out of the schools. The test was administered under usual conditions. Test papers were scored and the following results were obtained on the performance of these 12 children:

6 boys stood above 90th percentile.

2 boys stood between 80th and 90th percentile.

4 girls stood above 90th percentile.

These results very clearly show that Shankar's Creative Writing and Painting award winners scored very high on the present test, indicating thereby its high validity. Though

the number was small due to circumstances beyond our control, results were very useful and the uniqueness of the approach of the investigator to the problem of validity should merit due consideration.

(ii) TEACHERS RATING SCALE

Another criterion taken up as a reference for finding validity of the test is teacher-rating of these students on whom the creativity test was administered. A sample of 70 students was selected for the purpose. The students were rated by four teachers who taught them. Description of the rating scale[1] is as follows.

It is a sheet of paper which contained the definition of creativity and instructions to rate the student. The scale was divided into three points- highly creative, average creative and low creative. One box was marked under each point. The teacher was expected to mark a tick in either of the three boxes indicating thereby his estimate of the creativity of the student he was rating.

---

1 A specimen of Teachers Rating Scale is given in Appendix L.

Weightage was given to each rating. Highly Creative (HC) was given a score of three, Average Creative (AC) a score of two and Low Creative (LC) a score of one. As each student was being rated by four teachers, the maximum score obtained would be 12 and minimum 4.

The scores for each student were added. Correlation was computed between rating score and creativity test score. Again, the same following formula was made use of :

$$r = \frac{\Sigma XY - C_x C_y}{N \sigma_x \sigma_y}$$

$$= \frac{\frac{16}{70} - (.08 \times -.48)}{.79 \times 2.50}$$

$$= .13$$

where

r = Correlation between X and Y variable, viz. Teacher Rating and present test.

C_x = Correction on Teacher's Rating Score.

C_y = Correction on Creativity Test Score.

σ_x = Standard Deviation of Teacher's Rating Score.

σ_y = Standard Deviation of Creativity Test Score.

Σxy = Sum of products of scores on two variables.

N = Total Number of Sample.

The correlation thus obtained is positive but low. This is as was expected in view of halo effect. The teachers have broad notions about creativity. Their judgment is in fact an estimate determined by factors like scholastic achievement and perhaps by good behaviour of the students and not by the factors that are really involved in judging creativity. The fact, however, that correlation is positive and not negative indicates that the test is a good measure of creativity.

(iii) Present Test in its Relation with Intelligence

With the development of the concept of creativity, a need for evolving a separate test to measure it has been felt. This implies that a distinction has been drawn between intelligence and creativity and, therefore, between intelligence tests and creativity tests. For any test of creativity it is essential to see that it is not an intelligence test. In fact, a number of investigations have been carried out which indicate that intelligence and creativity are not identical, though positively related. The investigator had, therefore, to make sure that her test measured creativity and not intelligence. For this purpose, she administered a test of intelligence and found out correlation between intelligence test and the creativity

test that she had prepared. For this purpose Raven's Progressive Matrices (standard) was utilized.[1] The testing was done in two sittings. In one sitting the Intelligence test was administered and in another sitting, creativity test. Answers were scored. Correlation was found between the two scores. The same following formula was used :

$$r = \frac{\Sigma XY - C_n C_y}{\sigma_1 \sigma_2}$$

$$= \frac{\frac{63}{50} - .87 \times .02}{1.91 \times 2.46}$$

$$= .26$$

where

r = Correlation between X and Y variables, viz. intelligence and creativity.

C_x = Correction on Intelligence Test Score.

C_y = Correction on Creativity Test Score.

σ_x = Standard Deviation of Intelligence Test Score.

σ_y = Standard Deviation of Creativity Test Score.

ΣXY = Sum of Products of Scores on two Variables.

N = Total Number of Sample.

---

1 A Scoring Sheet is given in the Appendix M.

As seen above the correlation obtained is positive but low. This fact confirms that test measures creativity and not intelligence. The correlation is positive and this indicates that intelligence and creativity are definitely related.

We may recall here and summarize that for purposes of finding validity we took up the following reference criteria:

(1) Torrance Test of Creative Thinking against which the present test had a correlation of .72.

(2) Shankar's Painting and Writing Competition - the winners of which scored very high i.e. above 80 percentile on this test.

(3) Teachers rating scale against which the present test had a positive correlation.

(4) The fourth criterion which was administered to see if the test measures creativity more than intelligence indicates as shown above that the test is a measure more of creativity than for intelligence.

(iv) FACTOR ANALYSIS

It may be thus observed that the investigator took sufficient care to validate her test against the well established criteria and also showed that the test measured creativity and not intelligence. Normally, this should have been a sufficient measure for accepting the validity of the test as a whole as established. The totality of the test thus established, the investigator wished to examine if the factor constituting the test could also be validated so that the constituent factor as well as the test as a whole could be validated. The investigator therefore factor-analyzed the test performances, taking a sample of 100 students.

"Factor analysis is a method of analyzing a set of intercorrelated performances into as many independent variable factors as justify the labour of computation. Each factor is defined by the degree to which it participates in each of the various original performances."[1]

Factor analysis begins with a correlation matrix. A correlation matrix is a table of correlations with

---

1 Guilford,J.P.: Psychometric Methods, McGraw-Hill Book Co. Inc., New York, 1954, P.470.

columns and rows. It is symmetrical about its principal diagonal. The coefficients in the upper right portion are identical with those in the lower left portion. The nine variables considered here were as under :

(1) Scores on Fluency as measured by the test constructed by the investigator.

(2) Scores on Fluency as measured by Torrance Tests of creative thinking.

(3) Scores on Flexibility as measured by the test of the investigator.

(4) Scores on Flexibility as measured by Torrance tests of creative thinking.

(5) Scores on Originality as measured by the test of the investigator.

(6) Scores on Originality as measured by Torrance tests of creative thinking.

(7) Scores on Organization as measured by the test of the investigator.

(8) Scores on Imagination as measured by the test of the investigator.

(9) Scores on Richness as measured by the test of the investigator.

The data were analysed with the help of a computer. This was done mostly to avoid the errors inherent in working out mathematical calculations by hand and secondly to save the enormous amount of time that factor analysis would take if done by hand. The following tables 22-25 give the correlation coefficient matrix and other derivations.

Table 22: Showing Correlation Matrix of Nine Variables

Variables	Factors								
	1	2	3	4	5	6	7	8	9
1	(.99)	.88	.86	.70	.48	.44	.28	.26	.50
2	.88	(1.00)	.77	.72	.41	.47	.33	.16	.49
3	.86	.77	(.99)	.70	.67	.58	.25	.32	.50
4	.70	.72	.70	(.99)	.54	.51	.43	.29	.40
5	.48	.41	.67	.54	(.99)	.75	.33	.42	.42
6	.44	.47	.58	.51	.75	(.99)	.35	.44	.41
7	.28	.33	.25	.43	.33	.35	(.99)	.35	.60
8	.26	.16	.32	.29	.42	.44	.35	(.99)	.47
9	.50	.49	.50	.40	.42	.42	.60	.47	(.99)

Initially the data was fed to the computer for the purpose of extracting maximum numbers of factors. The axes was rotated nine times to start with. Following is the complete picture of principal components.

Table 23: Showing Principal Components of Nine Variables

Variables	Factors								
	1	2	3	4	5	6	7	8	9
1	.85	-.40	.13	.20	.02	-.01	-.16	-.03	.20
2	.82	-.43	.21	.03	.00	.23	-.04	-.17	-.14
3	.88	-.29	-.13	.13	.09	-.14	-.08	.24	-.11
4	.81	-.21	-.07	-.18	-.41	-.11	-.28	.02	.03
5	.75	*.18	-.48	-.20	.14	-.28	-.05	-.17	-.02
6	.74	*.21	.45	-.22	.08	.36	.04	.08	-.05
7	.55	*.49	-.50	-.39	-.10	-.02	-.21	.04	-.00
8	.51	.62	-.14	.50	-.29	.03	-.05	.03	-.03
9	.70	.35	.39	.17	.39	-.03	.24	.00	.02
Variance	55.46	14.35	10.41	6.78	5.00	3.28	2.44	1.39	.85

This table clearly shows a common factor running in all the variables which we may name as G factor. Factor 1 in this case can be designated as that G factor. This means that each variable is contributing something towards this G factor which we may call creativity.

To locate the specific factor contribution of each variable, factor loadings of each factor were found out. The axes was rotated nine times. The factor loadings indicate four main factor moving up. However, the picture is not very clear. The following table gives the factor loadings for nine variables.

Table 24: Showing Nine Factor Rotation

Variables	Factors								
	1	2	3	4	5	6	7	8	9
1	.93	.09	-.17	.11	.16	.09	.15	.01	.19
2	.89	.14	-,05	-.00	.17	.20	.21	-.08	-.25
3	.76	.02	-.36	.13	.18	.21	.19	.41	.02
4	.53	.22	-.20	.10	.07	.17	.77	.04	-.01
5	.25	.12	-.85	.18	.12	.35	.15	.04	.00
6	.25	.13	-.36	.21	.11	.85	.13	.03	-.01
7	.11	.93	-.09	.14	.24	.10	.13	.00	-.00
8	-.08	.14	-.14	.95	.17	.15	.06	.02	.00
9	.29	.32	-.13	.23	.85	.11	.06	.03	-.00
Variance	2.7	1.1	1.1	1.1	.92	1.01	.75	.18	.09

Column I gives a clear picture of one factor coming up. The rest of the columns have high loading for one variable only e.g. in column 2 variable 7, in column 3 variable 5, in column 5

variable 8, and so on. Practically all the variance is in the first four factors. This called for rotation of axes for four factors only. The table below shows all the four factors coming up distinctly.

Table 25: Showing Four Factors Rotation

Variables	Factors			
	1	2	3	4
1	.93	.09	-.17	.17
2	.91	.20	-.16	.01
3	.82	.02	-.42	.20
4	.70	.31	-.40	-.02
5	.30	.13	-.85	.20
6	.27	.16	-.84	.20
7	.12	.94	-.19	.12
8	.05	.16	-.28	.89
9	.40	.60	-.08	.52
Variance	3.21	1.45	1.95	1.23

Column 1 shows high loadings on first four variables, meaning thereby that there is something common in the four variables. The four variables constitute two fluency scores

(one score obtained on investigator's test and another on Torrance test) and two flexibility scores (one obtained on investigator's test and the other on Torrance test). Fluency and flexibility are highly related aspects. Fluency as defined earlier is the ability to give as many responses as he can, while flexibility is the ability to shift from one idea to another. In other words, flexibility is a measure of variation in responses. As the number of responses go on increasing, there is every chance that the responses would vary from one another. Not many relevant factors can be extracted from the same aspects. One has to shift the aspect or angle of perception to get more responses. Every new aspect will therefore bring in fresh response. This is how these two aspects are so closely related. Flexibility is in fact dependent on fluency. More of fluency means more of flexibility. More of fluency will never mean less of flexibility. The two are positively related. It is perhaps this reason that these two scores on both the tests show a high correlation with each other and a high factor loading has brought both the variables of the two tests under one factor. We, therefore, call this common factor as fluency.

Column 3 gives maximum loadings on two variables namely Originality I and Originality II (one obtained on investigator's test and another on Torrance's test). Both the variables measure the same ability and were therefore expected to get bracketed. There is however, a slight difference between the method of scoring originality on two tests. The score as obtained on investigator's test includes test score obtained by two methods of measurement. One is the content analysis method applied to two items and other is statistical method applied to the rest six items. The two scores added gives originality score. The score has high correlation with the score obtained on Torrance test of creative thinking which uses only one method i.e. statistical method. This common factor thus gets its natural name i.e. 'Originality'.

Column 2 gives maximum loadings on two variables i.e. Organization and Richness. Richness refers to the content of the material, while organization refers to the arrangement of the material. By content we mean the total number of ideas and their magnitude. The concept of organization gets meaning only when there is more than one idea. The greater the number of ideas, the bigger is the problem of their organization and better are the

chances of their arrangement and flow. The two concepts are therefore highly related. In fact organization is dependent on richness. We therefore, call this factor as richness.

Column 4 gives maximum loading on two variables, namely Imagination and Richness. The two concepts have much in common. The higher the imagination power, the greater are the number of ideas one can produce. In other words, higher the flight of ideas, wider becomes the view of the problem in hand. Wider perspective means that looking at more than one aspect of the situation and this leads one to get more ideas i.e. higher the imagination, higher is the richness. As the two aspects are highly related and seem to be measuring something common of the mental abilities, we call it as one factor and name it as 'Imagination'.

Thus, four factors in all have been identified, namely Fluency, Originality, Imagination and Richness.

The test when constructed was presumed to measure six factors out of which four factors have clearly been identified with the factor analytic method.

It has also been seen that the three variables measured by Investigator's test and Torance test have appeared as common factors.

These two considerations give a clear indication of the validity of the test. This establishes that the test measures what it was meant to measure.

The above discussion on interpretation of factors may be subjective, as understood by the present investigator, and opinions may differ. However, this is the best and utmost that the investigator could do and her approach, she feels, deserves meritorious consideration.
