CHAPTER - IX : RELIABILITY AND VALIDITY OF THE INVENTORY

(1) Reliability of the Inventory

Different methods are used to calculate the reliability of a test. These methods can be classified into two broad categories namely : (1) methods that depend upon the technique of correlations; (2) methods which make use of item statistics.

The following table gives an idea of the different methods under each of the two categories :

Methods to find Reliability

Methods depending upon correlation	Methods depending upon item statistics
1. Use of parallel forms	1. Kuder-Richardson Method
2. Test-retest method	2. Method of analysis of variance
3. Split-half method.	

Reliability of the present inventory

The reliability of the inventory has been measured by

(1) the test-retest method; (2) the split-half method and (3) the Kuder-Richardson method, (4) the Rulon Method; (5) The Flanagan Method.

(1) The Test-retest method

To apply this method, 47 students of the M.E.S.Teachers' Training College ~~students~~ were retested after an interval of about a month from the date of the first test. The two sets of scores in the two trials were then correlated and the co-efficient of reliability was found to be equal to 0.8930 (vide Table 36).

Date of first testing	29 - 9 - 1959
Date of second testing	27 - 10 - 1959
Mean score in the first test	102.8
Mean score in the second test	104.9

$$P.Er = \frac{.6745(1 - r^2)}{\sqrt{N - 1}} = \frac{.6745(1 - .893^2)}{\sqrt{47-1}} = 0.02$$

In the second trial the mean score has increased by 2.1. This may be due to the effects of memory, practice and familiarity or even chance. The reliability coefficient of 0.89 ± .02 is fairly high and this shows the reliability of the inventory.

(2) The Split-half method

In this method separate scores were derived by scoring

Table 36 - Reliability by the Test-Retest Method

Correlation between scores A in the First and Second Tests X ~~Scores B~~

Scores X in Second Test — y - variable	Scores IX in first test — x-variable: 60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139	Total
130-139							1	2	3
120-129					1	3	1	1	6
110-119					3	4	3		10
100-109				1	4	4			9
90-99				5	3	1			9
80-89		1	2	2					5
70-79			2						2
60-69		2							2
50-59	1								1
Total	1	3	4	8	11	12	5	3	47

M_x = 104.9 ; M_y = 102.8 ; $r_{x,y}$ = 0.8930 the odd-and-even numbered items. The correlation between these two scores was found to be 0.73 which gives the half test reliability (vide Table 37). From this half-test reliability the coefficient of reliability of the whole battery was calculated by applying Spearman-Brown formula :

$$r_{22} = \frac{2r_{11}}{1 + r_{11}}$$

where r_{22} is the coefficient of reliability of the whole test and r_{11} is the correlation between the two halves. Thus,

$$r_{22} = \frac{2 \times .73}{1 + .73} = 0.86.$$

(3) Kuder-Richardson Method

In this method, the test is split up into as many parts as there are items. Each part, therefore, contains an item. The number of correct responses to each item is then found out. This method differs from the split-half method in that it does not make use of correlation. This method can only be used if there is only one main variable or factor which the test is measuring. If more variables arø factors are involved, the reliability will be low. The present inventory involves only one variable, as can be seen at the end of the chapter. Therefore, the application of this method to find the reliability of the test

Table 37 - Reliability by the split-half method

Correlation between scores of odd and even items

Odd scores —— y-variable	Even scores —— x-variable													Total
	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	71-75	76-80	81-85	
81-85								1		1				2
76-80								1	1	2	2	2	1	9
71-75								1	4	4	2	3	1	15
66-70						1	4	3	9	13	9	3	3	45
61-65				1	2	5	11	13	20	19	7	2		80
56-60					3	12	16	23	23	9	9	2	1	98
51-55			1		1	18	19	25	9	13				86
46-50				1	18	15	22	20	2					78
41-45		1	2	9	17	10	4	3	2					48
36-40	1	1	3	3	14		2							24
31-35		2	3		2									7
26-30	1	3	2		2									8
Total	2	7	11	14	59	61	78	90	70	61	29	12	6	500

$M_x = 56.2$; $M_y = 44.0$; $r_{x,y} = 0.73$;

On using the Spearman-Brown Formula the stepped-up value is $r = 0.86$.

is quite justified.

The Kuder-Richardson formula is mathematicall represented as -

$$\gamma_{x} = \frac{n}{n-1} \times \frac{\sigma^2 - \sum pq}{\sigma^2}$$

where γ is the reliability coefficient of the whole test. n = number of items in the test; σ = standard deviation of the test scores; p = proportion of correct responses to each item; q = proportion of wrong responses to each item.

As the present inventory has been administered to a sample of 500 teachers, it will involve a tremendous amount of mathematical work to compute p and q. To minimise this, a sample of 100 answer sheets was selected at random from the 500 answer sheets. From the pile every 5th sheet was taken. The mean and the standard deviation of the scores were found out. Mean = 110.0 ; standard deviation = 19.0. These values do not differ significantly from the mean and standard deviation of scores of the 500 answer sheets. Therefore the sample of 100 answer sheets may be taken to be representative of the total sample of five hundred sheets. An item analysis chart of responses contained in the 100 answer sheets was prepared. From the chart, the number of correct responses to each of the 100 items was found out. The proportion of correct responses was calculated by the formula :

$$p = \frac{\text{total no. of correct responses to an item}}{\text{total number of cases}}$$

Table 38 gives the values of p and q for each item; q is found out by the formula q = 1-p. $r = 0.9554$

(4) Reliability by using Rulon Formula

Rulon has developed a simple formula for reliability of the total test scores. This formula involves the standard deviation of the differences between the scores on odd and even items. The standard deviation of the differences in the present case was computed using the same sample which was used to find the reliability by the K-R Method. The Rulon formula is $r = 1 - \frac{\sigma_d^2}{\sigma^2}$ where r = reliability of the test, σ_d = SD of the differences and σ = SD of the sample. The calculations are shown in Table 39.

Table 38 - Reliability by using the Kuder-Richardson Formula

Item No.	No. of correct responses	p proportion of correct responses	q = 1 - p Proportion of wrong responses	p x q
1	2	3	4	5
1	93	.93	.07	.0651
2	68	.68	.32	.2176
3	82	.82	.18	.1476
4	77	.77	.23	.1771
5	55	.55	.45	.2475
6	40	.40	.60	.2400
7	88	.88	.12	.1056
8	65	.65	.35	.2275
9	70	.70	.30	.2100
10	69	.69	.31	.2139
11	52	.52	.48	.2496
12	46	.46	.54	.2484
13	52	.52	.48	.2496
14	73	.73	.27	.1971
15	68	.68	.32	.2176
16	67	.67	.33	.2211
17	53	.53	.47	.2491
18	46	.46	.54	.2484
19	49	.49	.51	.2499
20	56	.56	.44	.2464
21	26	.26	.74	.1924
22	66	.66	.34	.2244
23	43	.43	.57	.2451
24	49	.49	.51	.2499
25	36	.36	.64	.2304
26	71	.71	.26	.2059
27	77	.77	.23	.1771
28	85	.35	.15	.1275
29	95	.95	.05	.0475
30	79	.79	.21	.1659
31	77	.77	.23	.1771
32	65	.65	.35	.2275
33	69	.69	.31	.2139
34	42	.42	.58	.2436
35	95	.95	.05	.0475
36	49	.49	.51	.2499
37	89	.89	.11	.0979
38	56	.56	.44	.2464
39	79	.79	.21	.1659
40	48	.48	.52	.2496
41	59	.59	.41	.2419
42	39	.39	.61	.2379
43	90	.90	.10	.0900
44	78	.78	.22	.1716
45	64	.64	.36	.2304
46	67	.67	.33	.2211
47	62	.62	.38	.2356
48	69	.69	.31	.2139
49	54	.54	.46	.2484
50	74	.74	.26	.1924
51	56	.56	.44	.2464
52	64	.64	.36	.2304
53	55	.55	.45	.2475
54	53	.53	.47	.2491
55	72	.72	.28	.2016
56	64	.64	.36	.2304
57	57	.57	.43	.2451
58	74	.74	.26	.1924
59	62	.62	.38	.2356
60	43	.43	.57	.2451
61	55	.55	.45	.2475
62	62	.62	.38	.2356
63	92	.92	.08	.0736
64	63	.63	.37	.2331
65	78	.78	.22	.1716
66	86	.86	.14	.1204
67	76	.76	.24	.1824
68	70	.70	.30	.2100
69	85	.85	.15	.1275
70	79	.79	.21	.1659
71	83	.83	.17	.1411
72	67	.67	.33	.2211
73	73	.73	.27	.1971
74	83	.83	.17	.1411
75	80	.80	.20	.1600
76	80	.80	.20	.1600
77	84	.84	.16	.1344
78	81	.81	.19	.1539
79	77	.77	.23	.1771

Table 38 (contd.)

1	2	3	4	5
80	67	.67	.33	.2211
81	66	.66	.34	.2244
82	85	.85	.15	.1275
83	82	.82	.18	.1476
84	84	.84	.16	.1344
85	80	.80	.20	.1600
86	89	.89	.11	.0979
87	75	.75	.25	.1875
88	80	.80	.20	.1600
89	84	.84	.16	.1344
90	74	.74	.26	.1924
91	73	.73	.27	.1971
92	67	.67	.33	.2211
93	56	.56	.44	.2464
94	82	.82	.18	.1476
95	77	.77	.23	.1776
96	82	.82	.18	.1476
97	89	.89	.11	.0979
98	70	.70	.30	.2100
99	56	.56	.44	.2464
100	79	.79	.21	.1659
			$\sum pq$ =	19.3270

$$\sum pq = 19.3270$$

$$\sigma = 19.0$$

$$r = \frac{n}{n-1} \times \frac{\sigma^2 - \sum pq}{\sigma^2}$$

$$= \frac{100}{99} \times \frac{19^2 - 19.3270}{19^2}$$

$$= 0.9554.$$

Table 39 - Frequency Table

Class interval	Frequency	x'	fx'	fx'^{2}
20 and above	2	7	14	98
18-19	2	6	12	72
16-17	3	5	15	75
14-15	4	4	16	64
12-13	4	3	12	36
10-11	5	2	10	11
8-9	11	1	11	11
6-7	13	0	-	-
4-5	16	-1	-16	16
2-3	28	-2	-56	112
0-1	12	-3	-36	108
	100		-18	$\sum fx'^2 = 612$

σ = SD of the sample = 19.0 $\qquad r = 1 - \frac{4.93^2}{19^2}$

σ_d = SD of the differences = 4.93

$$\therefore r = 1 - \frac{24.35}{361} = 1 - 0.0674 = 0.9330$$

(5) Reliability by using the Flanagan Formula[1]

Flanagan gives a formula parallel to Rulon's. It

1. Guilford, J.P. Psychometric Methods, Second Edn. 1954, pp 379-8

estimates the error variance as the sum of the variances of the two halves. This can be seen in the formula, where the sum $\sigma_1^2 + \sigma_2^2$ is used in place of Rulon's σ_d^2:

$$r = 2\left(1 - \frac{\sigma_1^2 + \sigma_2^2}{\sigma^2}\right)$$

where σ_1 and σ_2 are the standard deviations of the halves , r , the reliability of the whole test. As in the case of the Rulon formula, this provides an estimate of total score reliability.

$$\sigma_1^2 = 127.45 \; ; \quad \sigma_2^2 = 104.05$$

Applying the Flanagan formula to the data we have,

$$r = 2\left(1 - \frac{127.45 + 104.05}{20.2^2}\right)$$

$$= 2\left(1 - \frac{231.5}{408.04}\right)$$

$$= 2(1 - .567)$$

$$= .433 \times 2$$

$$= .87.$$

<u>Comparison of different menthods</u>

The reliability of the present test has been computed by various methods. Any single method used to determine the reliability of a test involves some fluctuations of the scores due to various factors. Table[1] 40 shows the factors that may be responsible for the fluctuations in the individual's performance on a test.

Table 40

Sources of variation	Test retest method	Split-half method	Rulon method	K-T formula	Flanagan method
1 Variations arising with measuring procure itself	✓	✓	✓	✓	✓
2 Changes in the individual from day-to-day	✓	x	x	x	x
3 Changes in the specific sample of task	x	✓	✓	✓	✓
4 Chances in the individual's speed of work	✓	x	x	x	x

It can be seen from table 40 that different factors are responsible for the variation in performance of an individual in a test. Reliability coefficients for the present inventory by various methods vary a little. This may be due to one or the other above factors.

1. Thorndike, E.Hagen : Measurement and Evaluation in Psychology and Education. John Wiley and Sons Inc, New York p. 131.

Interpretation of the different reliability coefficients

Table 41 gives the reliability coefficients of the present inventory by different methods.

Table 41

Method	Test retest	Split-half	Rulon	K.R. formula	Flanagan method
r	0.89	0.86	0.933	0.95	0.87

With the above coefficients of reliability can we say that the present inventory is satisfactory ? Some standard is necessary against which the reliability coefficients should be judged satisfactory or not. Kelley suggests that a test with a reliability as low as 0.5 is useful for determining the status of a group in some subject whereas a reliability of more than 0.9 is useful in a differentiating the status of an individual in a group in the same subject. The reliability coefficient of the present inventory by any method is not less than 0.85. Therefore, it can be confidently concluded that the inventory is sufficiently reliable.

Validity

The most important and difficult problem in the construction of tests is to devise methods to determine the validity of tests. A test is said to be valid if it measures what it is designed to measure. To discover whether or not a test actually functions in this way, two sets of measures are obviously needed; those of the test itself and of the think it is stacked against the measuring rod, as it were. The latter is known as the criterion. The correspondence between the ability measured by a test and the same ability measured by some other objective method gives the validity of the test. The degree of correspondence is shown by the coefficient of correlation (known as validity coefficient) obtained from accurate measures of the test and criterion. This problem of establishing the validity of a test has become very difficult because an independent criterion against which a test can be validdated is not available in most cases.

Generally, there have been two main approaches to the solution of this problem depending upon the nature of the criterion measure one selects for validation. If a test is validated against some internal criterion, the validity is known as 'internal validity'. If it is validated against some independent external criterion, we have what is known as 'external validity'. In the first approach to determine the internal validity, a test is made valid by definition. No external criterion is fixed and the scores of the test are not correlated with any criterion scores. Such an internal validity embraces four different aspects of a test namely : (a) curricular validity; (b) concept validity, or construct validity; (c) internal consistency, and (d) intercorrelationsof sub-tests (Factorial Validity).

The validity of the present inventory has been determined for three of the four aspects mentioned above.

(1) <u>Curricular validity of the present inventory</u>

The chief basis of the curricular validity of a test is the sampling of the items. The selection of the items should be such that they cover all the areas that are chosen for the test. As has been already said five important areas have been chosen and it is seen as far as possible that the items cover all these five areas. However, in constructing the present inventory the following steps were taken to ensure curricular validity.

(1) Analysis of the various tests that were available;

(2) Giving importance to all qualities according to frequency ;

(3) Scrutiny and criticism of the items by experts and experienced teachers.

(2) <u>Internal consistency of the inventory</u>

The internal consistency of the present inventory is secured by the item analysis procedure and the subsequent selection of the test items.

The internal consistency of a test is said to be high when each item correlates highly with the total test scores taken as the criterion scores. In the present inventory the methods used in item analysis have been to compute the ' biserial r ' by using the Flanagan chart as well as by calculation. Since the total score is simply the sum of the scores on individual items, it is apparent that the correlation between each item and the total score is an outcome of the correlatinns between the item and each of the other items of the inventory. According to Conrad‡ "Item-total-test-correlation serves as measure of the functional consistency between the various items of a test. If the item-total-test-correlation for a particular test item is high, then that item is highly consistent or 'homogeneous' with the other items of the test. If the item-total-test correlations of all the items are high, all the items are highly consistent with one another and the interval consistency or 'homogeniety' of the entire test is high.

No item selected for the final form of the inventory has an item total-test-correlation less than 0.25. Thorndike[1] writes : "An item-test correlation of 0.25 represents an outstanding validity".

It can, therefore, be concluded with confidence that the present inventory possesses an appreciably high "internal consistency".

---

1. Thorndike, R.L : "Peronnel Selection", John Wiley and Sons, Inc. New York, 1949. p 293.

(3) Intercorrelations of sub-tests

The intercorrelations of the present inventory are calculated by finding product-moment correlations between each sub-test score and the total score. The sub-tests comprised of five areas that were tested by a variety of items. The five areas that were tested and also the serial numbers of the items which tested these areas in the final form of the inventory are shown in Table 42.

Table 42

Areas	Serial numbers of the items in the final form of the inventory which are included in each of the areas.
1 Professional skill (S)	2, 6, 7, 8, 10, 34, 37, 38, 44, 48, 50, 52, 54, 56, 58, 62, 73, 86, 93, 96, 98, 99, 100.
2 Acquaintance with the principles of psychology (P)	3, 4, 11, 12, 13, 15, 16, 18, 20, 21, 23, 30, 32, 39, 41, 60, 65, 68, 69, 60, 72, 95, 94.
3 Ability for class management and administration (A)	17, 24, 25, 31, 40, 42, 57, 61, 63, 64, 74, 75, 78, 79, 80, 81, 85, 87, 91, 97.
4 Relationship with others including authorities, colleagues, pupils and parents (R)	5, 9, 22, 26, 28, 29, 35, 43, 46, 47, 53, 71, 82, 90, 92.
5 Individual qualities (I)	1, 14, 19, 27, 33, 36, 45, 49, 51, 55, 59, 66, 67, 76, 77, 83, 84, 88, 89.

The intercorrelations between the five sub-tests are also calculated.

Table 43 - Correlation coefficients between sub-tests and the total test

Sub-test	1	2	3	4	5
	0.8638	0.7723	0.7401	0.7240	0.7462

Table 44 - Correlation Matrix of the sub-tests

	1	2	3	4	5
1	-	0.5126	0.5483	0.6365	0.6842
2	-	-	0.5165	0.4854	0.4427
3	-	-	-	0.4063	0.4434
4	-	-	-	-	0.4687
5	-	-	-	-	-

Both the tables 43 and 44 reveal fairly high positive correlations between sub-tests and the total and also between the sub-tests. The positive high correlations indicate the overlap of sub-tests on the another. It may be said that all sub-tests measure mostly one and the same factor. Also, in table 43 shows that all values of the correlations are more than 0.72, which means that all the sub-tests are in good agreement with the whole battery. The good teamwork of the sub-tests enhances the validity of the whole inventory.

External validity and the present inventory

When the criterian against which a test is validated is some external independent one, we get the external validity of the test. The external validity includes the following varieties :

(a) Congruent validity
(b) Concurrent validity
(c) Predictive validity.

If the scores of the inventory are correlated with scores of any previously standardised teacher-efficiency test, the correlation coefficient is called 'congruent validity'. As no tests of teacher aptitude or efficiency have been standardised with respect to the primary school teachers of Mysore State, the congruent validity of the test cannot be found out.

(b) Concurrent validity has been determined taking three external criteria.

(1) The terminal examination marks of pupil teachers

The terminal examination marks of the 47 pupil-teachers of the M. E. S. Teachers' College, Malleswaram, were used, as an external criterion to calculate the concurrent validity. The correlation between the examination marks and their scores on the inventory was 0.66. This is faily high (vide table 45).

Table 45 – Validity with respect to Examination Marks

Coefficient of correlation between the inventory scores and examination Marks

Inventory Score – y-variable	Examination marks —— x-variable								Total
	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	
130–139					1	1		1	3
120–129						1	4	1	6
110–119				1	4	4	2	3	10
100–109				2	4	2	1		9
90–99			4	1		2	2		9
80–89			1	2	1				4
70–79	1	1			1				3
60–69				1	1				2
50–59	1								1
Total	2	1	5	7	8	10	9	5	47

$M_x = 53.5$; $M_y = 102.8$; $r_{x,y} = 0.66$

(2) Supervisor's ratings

The ratings of the testees by experienced supervisors can constitute a valid and reliable criterian. Hence, the ratings of supervisors were taken as an external criterian for a group of pupil-teachers of Hospet Training School. Two teachers of the Hospet Training School were requested to rate the pupil-teachers on a five-point scale as very good, good, average, bad and very bad, bearing in mind their class-room teaching also. These two teachers had seen a number of lessons of those pupil-teachers and had a close contact with them. Hence, they were able to rate them properly. These ratings were reduced to percentage basis as follows :

Those who were rated :

as very good were given	scores of	60 % and above
as good were given	"	55 %
as average were given	"	50 %
as bad were given	"	25 %
as very bad were given	"	20 %

The correlation between the test scores and the percentages was calculated and the correlation was found to be 0.31. In view of the variability of the teachers' estimate, 0.31 may be said to be fairly good correlation (vide Table 46).

Table 46 - Validity with respect to supervisor's rating

Correlation between inventory scores and supervisor's rating

Supervisor's ratings y-variable	Inventory Scores — x-variable 50-59	60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139	Total
60-64							1		1	2
55-59					1	4	1	4	3	13
50-54			1		2	4				7
45-49	1			1	3	3	2	1	1	12
40-44					1	1	2	4		8
35-39				2	2					4
30-34						1				1
Total	1	0	1	3	9	13	6	9	5	47

$M_x = 107.9$; $M_y = 49.1$; $r_{x,y} = 0.31$

(3) Children's evaluation of teachers

The third external criterion used was the pupil evaluation of the teacher. In order to do this, a questionnaire with 20 statements was prepared (as shown in Appendix P). The pupils had to read the statement and decide how often the teacher about whom he would be writing, did what was asked in a particular statement. He had to indicate his answer by underlining "always, usually, sometimes or never" for each of the statements. These statements were supposed to elicit those qualities which the inventory tried to assess. Ratings were obtained from at least 25 pupils for one teacher and thus about 90 teachers were rated by a vast number of children. This rating scale was administered by the tester with no-one else in the room except the pupils. Assurance was given to the pupils that no one other than the tester would see the ratings, that their answer sheets would remain anonymous and that their promotions would not be affected. Pupils' ratings were obtained only with regard to teachers who taught the Middle Third and Fourth year classes since it was thought that the lower class pupils would find it difficult to rate a teacher. Hence only the 90 teachers teaching the Middle III and IV year classes were rated by the pupils. But the heads of some schools did not like their pupils to rate their teachers. Hence, all the teachers who taught the higher classes could not be rated by pupils. About 2700 children rated their teachers. Some of the children who could not follow the instructions were not able to rate the teachers properly. A few such ratings were discarded. Then some of the teachers had ratings from less than 20 pupils. Only the ratings for a teacher by at least 20 to 25 pupils were retained.

Administration of the questionnaire

When the children were seated in a class-room, the questionnaire was distributed and the following instructions were given to them :

> " Here you find some statements about your teacher. Read each statement and decide how often your teacher does what is asked about. Underline the word that shows how often he/she does it. Please answer the statements honestly. You are not asked to write your name on the paper. None of the teachers or the headmaster will ever see this paper or know how you have marked the statements. While marking your answer, think of the teacher whose name appears below ".

The pupils were also asked to write the name of the teacher in the space left for the same. When all the students had finished answering the questionnaire, the answer sheets were collected and the same procedure was followed with the next batch. The pupils enjoyed the statements and answered them honestly. A few words where meanings were a bit difficult for a few pupils were explained. Meanings of all difficult words were given to all the batches uniformly, so that there was no room for guessing the meanings. The pupils were able to understand and follow the instructions. While the pupils were answering, care was taken to see that they did not copy each other's answers or consult each other. The average score of the 20 to 25 pupils for each of the teacher was calculated by alloting arbitrary values of 3, 2, 1 and 0 for "always, usually, sometimes and never" respectively. The correlation between these scores and the scores obtained by these teachers on the inventory was computed. The correlation coefficient by the product-moment method was found to be equal to 0.46. (Vide Table 47). This definitely shows that the pupils rate their teachers better than supervisors.

(c) Predictive Validity

By 'Predictive Validity' is meant the efficiency with which the test scores can predict the future successes or failures of students in the particular trait measured by the test. The most common means of checking predictive validity is by correlating test scores with a subsequent criterian measure. Concurrent validity and predictive validity are quite similar except for the time at which the criterion is obtained. Also, a test having concurrent validity may not have predictive validity.

Table 47 - Validity with respect to children's evaluation

Correlation between inventory scores and children's evaluation of teachers

Children's evaluation — y-variable	Inventory scores — x variable										Total
	60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	150-159	
23-24									1		1
21-22						2	2	1	1		6
19-20						1	2	3		1	7
17-18					5	4	2	3	1	1	16
15-16					3	3	5	2	2	2	17
13-14	1		3	2	5	3	3	2	3		22
11-12		1		2	1	4	1				9
9-10				1	2	2	1				6
7-8				2		1					3
5-6		1			2						3
Total	1	2	3	7	18	20	16	11	8	4	90

$M_x = 117.2$; $M_y = 14.8$; $r_{x,y} = 0.46$

Any inventory should have a fairly high predictive validity if it should be used for selection purposes. In order to find out how well this inventory predicts the efficiency of the teachers, the correlation between the inventory scores and the marks obtained by the pupil-teachers of the M. E. S. Teachers' College, Bangalore and Teacher Training School, Hospet, in their final public examination was calculated. The correlation was found to be equal to 0.56 (vide Table 48). This is fairly high. Hence it can be said that the inventory has fairly high predictive validity. [Permission was secured by the department to utilise the public examination marks for experimental purposes (vide Appendix Q).]

In brief, all possible efforts have been made to establish the reliability and validity of the inventory, though some efforts might not be very satisfactory methodologically. Yet, on the whole, all the values obtained show that the inventory is reliable and valid for all practical purposes.

Factor Analysis

Factor analysis has been applied to aptitude, attitude and personality variables. Harrell[1] attempted to study the tests of mechanical ability by applying this technique. In

1. Harrell, T.W. : "A factor analysis of mechanical ability tests". Psychometrika, 1940, 5, pp 17-33.

Table 48 - Predictive Validity

Correlation between inventory scores and public examinations marks

Inventory scores (y-variable)	Public Examination Marks x-variable 130-149	150-169	170-189	190-209	210-229	230-249	250-269	Total
130-139				1	4		1	6
120-129				4	3	1		8
110-119				4	3	1	1	9
100-109				5	4	1		10
90-99	1	1	2	1	4			9
80-89			3	2				5
70-79	1							1
60-69			1					1
50-59			1					1
Total	2	1	7	17	18	3	2	50

$M_x = 205.5$; $M_y = 107.0$; $\sigma_x = 24.0$; $\sigma_y = 18.4$

$r = 0.56$

his study the factor analysis was made by a multiple factor method since the absence of a general factor is suggested by the original correlations. Hellfritzsch[1] made a study of teaching abilities by the application of factor analysis. In this study also an attempt has been made to investigate the factors the inventory measures and the extent they are measured by applying factor analysis. For this purpose, the internal correlations of the five sub-tests have been calculated from a sample of 100 answer-sheets selected on a random basis. The factorial analysis of the tests is done by applying Thurstone's centroid method and is verified by Spearman's formula. This method has been tried since the correlations between the sub-tests and the inventory and the intercorrelations of the sub-tests suggest the presence of a general or common factor running through the inventory.

Table 49 - Original correlation matrix
(N = 100)

Test	1	2	3	4	5
1	-	0.5126	0.5483	0.6365	0.6842
2	0.5126	-	0.5165	0.4854	0.4427
3	0.5483	0.5165	-	0.4063	0.4434
4	0.6365	0.4854	0.4063	-	0.4687
5	0.6842	0.4427	0.4434	0.4687	-

To apply the Centroid Method of factoring, the highest correlation coefficient in a column is fixed up as the

1. Hellfritzsch, A.G. :A factor analysis ofteaching abilities" Journal of Experimental Education, 1945, XIV, p 166.

communality in that column. In the following matrix the numbers in the perantheses are the communalities :

Table 50

Test	1	2	3	4	5
1	(0.6842)	0.5126	0.5483	0.6365	0.6842
2	0.5126	(0.5165)	0.5165	0.4854	0.4427
3	0.5483	0.5165	(0.5483)	0.4063	0.4434
4	0.6365	0.4854	0.4063	(0.6365)	0.4687
5	0.6842	0.4427	0.4434	0.4687	(0.6842)
Total	3.0658	2.4737	2.4628	2.6334	2.7232
	=	13.3589	= 3.6550^2		
First factor Loadings	0.8388	0.6768	0.6738	0.7205	0.7451

The correlations of each column in Table 50 are totalled and the totals are summed up (13.3589). The square root of this grand total is found out (3.6550). All the column totals are then divided by this square root giving the loadings of the first factor. From these loadings the first factor matrix (Table 51) is obtained.

Table 51 - First Factor Matrix

First factor matrix	0.8388	0.6768	0.6738	0.7205	0.7451
0.8388	(0.7036)	0.5677	0.5652	0.6044	0.6250
0.6768	0.5677	(0.4581)	0.4560	0.4876	0.5043
0.6738	0.5652	0.4560	(0.4540)	0.4855	0.5020
0.7205	0.6044	0.4876	0.4855	(0.5191)	0.5368
0.7451	0.6250	0.5043	0.5020	0.5368	(0.5552)

The above matrix is then subtracted from the original correlation matrix to find the residues.

Table 52 - First Residual Matric

	1	2	3	4	5
1	(-0.0194)	-0.0551	-0.0169	0.0321	0.0592
2	-0.0551	(0.0584)	0.0605	-0.0022	-0.0616
3	-0.0169	0.0605	(0.0943)	-0.0792	-0.0586
4	0.0321	-0.0022	-0.0792	(0.1174)	-0.0681
5	0.0592	-0.0616	-0.0586	-0.0681	(0.1290)

The algebraic sums of elements in every column of the residual matrix are :

-0.0001 0.0000 0.0001 0.0000 -0.0001

which checks to zero.

Extraction of second centroid factor

Before second factor can be extracted some test vectors ~~shall~~ have to be reflected. Before doing this, we write the residual matrix fixing the new communalities. This is done by removing diagonal elements in the residual matrix and substituting in their places the highest elements from the columns. This gives the following matrix :

Table 53

	1	2	3	4	5
1	(0.0592)	-0.0551	-0.0169	0.0321	0.0592
2	-0.0551	(0.0616)	0.0605	-0.0022	-0.0616
3	-0.0169	0.0605	(0.0792)	-0.0792	-0.0586
4	0.0321	-0.0022	-0.0792	(0.0792)	-0.0681
5	0.0592	-0.0616	-0.0586	-0.0681	(0.0681)

The sums disregarding the signs of the columns are :

0.2225 0.2410 0.2944 0.2608 0.3156

The highest sum is 0.3156 of the column 5. We reflect test vectors 2, 3 and 4 so that all elements of column 5 are positive. The general policy is to reflect one test vector at a time, note the results and then reflect the second and so on. The matrix after the reflection of test vectors 2, 3 and 4 is given in Table 54.

Table 54

	1	2	3	4	5
1	0.0592	0.0551	0.0169	-0.0321	0.0592
2	0.0551	0.0616	0.0605	-0.0022	0.0616
3	0.0169	0.0605	0.0792	-0.0792	0.0586
4	-0.0321	-0.0022	-0.0792	0.0792	0.0681
5	0.0592	0.0616	0.0586	0.0681	0.0681
Total	0.1583	0.2366	0.1360	0.0338	0.3156
		=	0.8803 = 0.9382^2		

Dividing every column sum by 0.9382 we get the second factor loadings as :

0.1687 0.2522 0.1450 0.0360 0.3364

Before calculating the second factor loadings, Humphrey's test of Residues was applied to find out whether the second factor is present or not.

The product of the highest factor loadings is 0.3364 x 0,2522 = 0.0848. This product is less than $\frac{2}{\sqrt{N}}$, where N is the sample size which is 100 in this case.

$$\frac{2}{\sqrt{100}} = \frac{2}{10} = 0.2$$

0.0848 is less than 0.2.

Since the product of the two highest factor loadings of the second factor is less than $2/\sqrt{N}$, the second factor is not present. Hence, only one factor is present and its loadings are :

0.8388 0.6768 0.6738 0.7205 and 0.7451

The first factor loadings of tests indicate their saturation with the general factor. These saturations are further checked by Spearman's formula.

Spearman's method of determining the saturations of the test

According to Spearman $r_{ug} = \frac{A^2 - A'}{T - 2A}$ where A is the sum of correlations between the test U and every other test, A' is the sum of the squares of these correlations and T is the total of all correlations in the table.

Step I (Calculation of A^2) :

Table 55

Test	1	2	3	4	5	A	A^2
1	-	0.5126	0.5483	0.6365	0.6842	2.3816	5.6720
2	0.5126	-	0.5165	0.4854	0.4427	1.9572	3.830
3	0.5483	0.5165	-	0.4063	0.4434	1.9145	3.663
4	0.6365	0.4854	0.4063	-	0.4687	1.9969	3.990
5	0.6842	0.4427	0.4434	0.4687	-	2.0390	4.158

$T = \sum A = 10.2892$

Step II (Calculation of A')

Table 56

Test	1	2	3	4	5	A'
1	-	0.2628	0.3006	0.4052	0.4681	1.4367
2	0.2628	-	0.2667	0.2356	0.1960	0.9611
3	0.3006	0.2667	-	0.1650	0.1968	0.9291
4	0.4052	0.2356	0.1650	-	0.2212	1.0270
5	0.4681	0.1960	0.1968	0.2212	-	1.0821

Step III (Calculation of the saturation of the general factor)

Table 57

Test No.	A^2	A'	A^2-A'	2A	T-2A	$\frac{A^2-A'}{T-2A}$	Saturation of the gen. factor
1	5.6720	1.4367	4.2353	4.7632	5.5260	0.7653	0.8754
2	3.8300	0.9611	2.6689	3.9144	6.3748	0.4188	0.6571
3	3.6630	0.9291	2.7339	3.8290	6.4602	0.4223	0.6506
4	3.9900	1.0270	2.9630	3.9938	6.2954	0.4707	0.6860
5	4.1580	1.0821	3.0759	4.0780	6.2112	0.4952	0.7037

Table 58 gives the saturations of the sub-tests calculated by Spearman's formula and their comparison with those calculated by Thurstone's centroid method. They are also compared with the correlations of the tests with the whole battery.

Table 58 - General factor - Saturations of the tests and their comparison with correlations of the tests with the whole inventory

	Correlations of the sub-tests with the inventory	General factor saturation		Order of the test as per the saturation of the general factor by	
		Spearman's formula	Thurstone's centroid method	Spearman's formula	Thurstone's formula
1 Professional skill and interest	0.8638	0.8754	0.8388	1	1
2 Acquaintance with the principles of psychology	0.7723	0.6571	0.6768	4	4
3 Ability for class management and organisation	0.7401	0.6506	0.6738	5	5
4 Relationship with others including authorities colleagues, pupils and parents	0.7240	0.6869	0.7205	3	3
5 Individual qualities	0.7462	0.7037	0.7451	2	2
	3.8464	3.5728	3.6550		

It is seen from table 58 that the saturations of the tests calculated by both the methods are almost the same. Moreover, they are quite high. This shows that the inventory measures a certain factor and it is a good measure of that factor. This factor may be called as 'Teacher-efficiency'. The order of the sub-tests in respect of saturation of 'teacher-efficiency' shows that the first area i.e. professional skill and interest is the best measure of teacher-efficiency. There is also a close resemblance of the 'correlation of the sub-tests with the inventory with the saturations of the general or common factor. This indicates that the total score on the inventory obtained by a testee may be taken as representing his teacher-efficiency.

:: :: ::