CH_APTER V

FIRST TRY OUT OF THE TEST

Objectives

The main objectives of the First Try out test are as follows:

1. To find out gross deficiencies in the test items such as to correct answers, ambiguous words etc.

2. To find out the difficulties of the pupils.

3. To determine the appropriate time limits for the tests

Procedure for the First Try Out

Sample

First Try out testing was carried out at various city schools. There are about 10 high schools in Srinagar. 50 students, irrespective of sex, were given mimeographic paper sheets for sub-test at a time. Each group to be studied was given the tests. They were selected from five schools situated in different localities. The following table shows the detail of the sample from Srinagar.

TABLE

School-wise, Sex-wise and Age-wise Distribution
of the Sample of the First Try Out of the
Test

Sr. No.	Name of the School	Age	Boys	Girls	Total
1.	S.P.H.S.School	12	50	-	50
2.	M.P.H.S.School	13	50	-	50
3.	V.G.H.S.School	14	-	50	50
4.	Islamia H.School	16	50	-	50
5.	G.H.S. Nawakadul	16	-	50	50
	Total		150	100	250

Administering the Test and

Collecting the Data

All the sub-tests were duplicated separately and
on sub-test was administered at a time to a group of
fifty pupils. All the pupils had to start together,
and work to the completion of the sheets. No time
limit was imposed. They were asked to return the
sheets to the Instructor on completion of their work.
The Instructor would note down the time taken by
each pupil.

The second sub-test was to be distributed when
all had completed the first sub-test. Similar
procedure was followed while administering the
remaining test to one group in two days.

<u>The Procedure</u>

Here is the procedure for the test in a spick and span manner.

1. The pupils were to be arranged in parallel rows and seated at convenient distances.

2. Each pupil could avail himself or herself of as much time as he or she required.

3. The answering process was arranged to be simultaneous, the pupils had been directed to start writing at the same time.

4. Time taken by each pupil was recorded on the paper sheets.

<u>Administering the First Try Out</u>

The first Try out Testing was carried out at M.P. H.S. Srinagar, Nawakadal Boys High School, Hindu High School, Srinagar and Vasant Girls High School, Srinagar in the month of June, 1962. The following table shows the number of pupils tested in each school.

T A B L E 2

Table showing the Age, Sex, Grade of the Pupils
Tested in the First Try Out Testing at
Srinagar

Sr. No.	Name of the School	Age	Boys	Girls	Total
1.	Vasant Girls High School	12	-	50	50
2.	Islamia High School	13	50	-	50
3.	M.P. High School	14	50	-	50
4.	Nawakadal Boys High School	15	50	-	50

5. Hindu High School
       Srinagar   16   50   -   50

          Total   200 50 250

The following table gives the idea of correct responses, incorrect responses of each item and their respective places according to the new order if accepted or rejected.

TABLE 3

Item No.	No. of correct Responses	No. of Incorrect Responses	No. of Unattempted	Total	New order if accepted or rejected
1	2	3	4	5	6
Sub-test Number I – Opposites					
1	170	75	5	250	1
2	115	112	23	250	15
3	145	92	13	250	4
4	133	100	17	250	6
5	120	120	10	250	14
6	147	95	8	250	3
7	127	107	16	250	11
8	130	105	15	250	9
9	62	170	18	250	Rejected
10	95	132	23	250	Rejected
11	110	127	13	250	9
12	35	205	10	250	Rejected
13	130	112	8	250	10
14	110	122	18	250	16
15	110	120	20	250	17
16	97	135	18	250	Rejected
17	120	122	8	250	12

(Contd.)

1	2	3	4	5	6
18	170	67	13	250	Rejected
19	125	117	8	250	19
20	132	92	26	250	7
21	95	137	18	250	Rejected
22	130	110	10	250	8
23	92	145	13	250	Rejected
24	80	142	8	250	"
25	100	127	23	250	20
26	110	120	20	250	18
27	132	110	8	250	5
28	92	147	11	250	Rejected
29	80	155	15	250	"
30	100	127	23	250	21
31	99	121	30	250	22
32	98	120	32	250	23
33	95	118	37	250	24
34	92	120	38	250	25
35	42	192	16	250	Rejected
36	88	127	35	250	27
37	64	176	10	250	Rejected
38	90	124	36	250	26
39	86	126	38	250	28
40	65	177	8	250	Rejected

1	2	3	4	5	6
		Sub-test Number II. SIMILARITIES			
1	147	92	11	250	9
2	115	130	5	250	20
3	160	80	10	250	3
4	155	85	10	250	4
5	85	160	5	250	Rejected
6	82	165	3	250	"
7	135	102	13	250	13
8	147	95	8	250	10
9	157	85	8	250	5
10	175	45	30	250	1
11	155	90	5	250	6
12	115	130	5	250	22
13	157	87	6	250	6
14	162	85	3	250	2
15	135	97	18	250	14
16	157	87	6	250	7
17	117	130	3	250	19
18	115	127	8	250	21
19	25	187	38	250	Rejected
20	120	125	5	250	17
21	55	180	15	250	Rejected
22	42	202	6	250	"
23	105	135	10	250	23
24	137	105	8	250	12
25	125	120	5	250	16
26	20	175	55	250	Rejected

Contd.

1	2	3	4	5	6
27	100	140	10	250	Rejected 24
28	107	137	6	250	" Rejected
29	100	140	10	250	25
30	97	137	16	250	26
31	127	116	8	250	15
32	145	95	10	250	15
33	120	120	10	250	11
34	47	187	16	250	Rejectedm
35	65	150	35	250	"
36	42	66	142	250	"
37	95	130	25	250	27
38	92	134	24	250	29
39	95	127	28	250	28
40	55	112	83	250	Rejected

Sub Test Number III. CLASSIFICATION

1	70	170	10	250	16
2	80	145	25	250	14
3	97	110	43	250	12
4	77	117	55	250	15
5	97	72	81	250	13
6	137	73	40	250	01
7	135	72	43	250	2
8	135	85	30	250	3
9	37	192	21	250	Rejected
10	55	175	20	250	9
11	85	135	30	250	5
12	33	192	25	250	Rejected
13	60	152	38	250	10

(Contd.)

1	2	3	4	5	6
14	90	45	115	250	4
15	67	145	38	250	7
16	80	135	35	250	6
17	33	165	52	250	Rejected
18	65	140	45	250	8
19	35	187	28	250	17
20	35	182	33	250	18
21	33	167	50	250	19
22	42	157	51	250	20
23	55	150	45	250	21
24	60	150	40	250	11
25	52	138	60	250	22
26	50	140	60	250	23
27	51	130	69	250	24
28	49	125	76	250	25
29	48	132	70	250	27
30	48	130	72	250	26

Sub-Test Number IV. ANALOGIES

1	192	36	22	250	1
2	92	125	33	250	15
3	122	102	26	250	10
4	182	50	18	250	2
5	112	130	8	250	11
6	30	187	33	250	Rejected
7	50	175	25	250	"

(Contd.)

1	2	3	4	5	6
8	55	180	15	250	Rejected
9	167	65	18	250	5
10	160	80	10	250	7
11	55	192	3	250	18
12	60	25	165	250	Rejected
13	45	85	120	250	"
14	61	52	137	250	17
15	55	112	83	250	Rejected
16	172	62	16	250	3
17	100	117	33	250	13
18	110	122	18	250	12
19	160	57	33	250	6
20	85	137	28	250	16
21	147	65	38	250	8
22	132	87	31	250	4
25	47	197	6	250	Rejected
26	52	192	6	250	19
27	116	104	30	250	11
28	80	132	38	250	20
29	80	160	10	250	Rejected
30	50	140	60	250	21
31	48	130	72	250	22
32	40	60	150	250	Rejected
33	48	184	18	250	23
34	47	141	62	250	24

(Contd.)

1	2	3	4	5	6
35	47	155	48	250	25
36	45	160	45	250	26
37	45	157	48	250	27
38	43	149	58	250	28
39	42	147	61	250	30
40	43	150	57	250	29

Sub-test Number V. PROBLEMS

1	2	3	4	5	6
1	197	30	23	250	1
2	122	102	26	250	10
3	160	80	10	250	7
4	55	192	3	250	Rejected
5	92	125	33	250	15
6	182	50	18	250	2
7	30	187	33	250	Rejected
8	57	67	126	250	"
9	45	85	120	250	"
10	50	45	155	250	"
11	167	65	18	250	5
12	62	146	52	250	17
13	112	130	8	250	11
14	172	58	20	250	3
15	52	53	145	250	Rejected
16	37	70	143	250	"
17	42	172	16	250	"
18	100	117	33	250	13
19	110	122	18	250	12
20	160	57	33	250	6

(Contd.)

1	2	3	4	5	6
21	47	197	6	250	Rejected
22	49	190	11	250	"
23	52	182	16	250	"
24	85	137	28	250	16
25	147	65	38	250	8
26	47	190	13	250	Rejected
27	132	87	31	250	9
28	170	65	15	250	4
29	115	105	30	250	14
30	40	197	13	250	Rejected

Sub-test Number VI. <u>NUMBER SERIES</u>

1	2	3	4	5	6
1	137	72	41	250	1
2	135	85	30	250	3
3	60	152	38	250	9
4	85	135	30	250	5
5	55	175	20	250	10
6	33	192	25	250	Rejected
7	90	45	115	250	4
8	67	145	38	250	7
9	80	135	35	250	6
10	33	165	52	250	Rejected
11	65	140	45	250	8
12	62	180	8	250	Rejected
13	35	187	28	250	16
14	35	182	33	250	17
15	27	185	38	250	21
16	55	150	45	250	19

Contd.

1	2	3	4	5	6
17	25	190	35	250	22
18	42	157	51	250	20
19	23	187	40	250	22
20	135	72	43	250	2
21	23	187	40	250	24
22	97	72	81	250	14
23	97	110	43	250	15
24	80	145	25	250	13
25	77	117	55	250	12
26	33	167	50	250	18
27	57	173	20	250	11

Sub-test Number VII. <u>JUMBLED SENTENCES</u>

1	2	3	4	5	6
1	165	82	3	250	1
2	130	120	-	250	13
3	139	112	1	250	10
4	140	110	-	250	6
5	135	107	8	250	7
6	135	112	3	250	8
7	155	92	3	250	4
8	157	90	3	250	3
9	110	137	3	250	Rejected
10	122	125	3	250	, ,
11	117	127	6	250	, ,
12	130	117	3	250	14
13	150	95	5	250	5
14	160	87	3	250	2
15	127	120	3	250	16

(Contd.)

1	2	3	4	5	6	
16	117	127	6	250	Rejected	
17	135	112	3	250	12	
18	130	115	5	250	15	
19	140	97	13	250	8	
20	140	95	15	250	9	
21	125	120	5	250	17	
22	124	118	8	250	18	
23	123	116	11	250	19	
24	62	170	18	250	Rejected	
25	92	145	13	250	"	
26	122	119	9	250	20	
27	120	115	15	250	23	
28	121	118	11	250	22	
29	119	117	24	250	24	
30	122	117	21	250	21	

The percentage of correct responses of the
items retained, i.e. 179 were also found out and
it was seen that they also tallied with the correct
responses in the previous table and are also in
desending order

T A B L E 4

Table 4 showing the Number of Correct Responses
and their Percentage in the First Try Out
of 179 Items

Item No:	Correct Responses	Percentage
1	2	3
Sub- Test Number I - Opposites		
1	170	68.0
2	115	46.0
3	145	58.0
4	133	53.2
5	120	48.0
6	147	58.8
7	127	50.8
8	130	52.0
9	110	44.0
10	130	52.0
11	110	44.0
12	110	44.0
13	120	48.0
14	170	68.0
15	125	50.0
16	132	52.8
17	130	52.0

(Contd.)

1	2	3
18	100	40.0
19	110	44.0
20	132	52.9
21	100	40.0
22	99	39.6
23	98	39.2
24	95	38.0
25	92	36.8
26	90	36.0
27	88	35.2
28	86	34.4

Sub-test Number II. <u>SIMILARITIES</u>

1	147	58.8
2	115	46.0
3	160	64.0
4	155	62.0
5	135	54.0
6	147	58.8
7	157	62.8
8	175	70.0
9	162	64.8
10	135	54.0
11	155	62.0
12	115	46.0
13	157	62.8
14	157	62.8
15	115	46.0

(Contd.)

1	2	3
16	117	46.8
17	120	48.0
18	105	42.0
19	137	54.8
20	125	50
21	100	40.0
22	100	40.0
23	97	38.8
24	127	50.8
25	145	58.0
26	120	48.0
27	95	38.0
28	92	36.8
29	95	38.0

Sub-test No:III. CLASSIFICATION

1	70	28.0
2	80	32.0
3	97	38.8
4	77	30.8
5	97	38.8
6	137	54.8
7	135	54.0
8	135	54.0
9	55	22.0
10	85	34.0
11	60	24.0

(Contd.)

1	2	3
12	90	36.0
13	67	26.8
14	65	26.0
15	80	32.0
16	35	14.0
17	35	14.0
18	33	13.8
19	42	16.8
20	55	22.0
21	60	24.0
22	52	20.8
23	50	20.0
24	51	20.4
25	49	19.6
26	48	19.2
27	48	19.2

Sub-test No.IV. ANALOGIES

1	198	76.8
2	92	36.8
3	122	48.8
4	182	72.8
5	112	44.8
6	117	46.8
7	160	64.0
8	55	22.0
9	61	24.4
10	172	68.8

(Contd.)

1	2	3
11.	100	40.0
12	110	42.0
13	160	64.0
14	85	34.0
15	147	58.8
16	132	52.8
17	115	46.0
18	176	68.1
19	140	56.0
20	80	32.0
21	50	20.0
22	48	19.2
23	48	19.2
24	47	18.8
25	47	18.8
26	45	18,1
27	45	18.1
28	43	17.2
29	43	17.2
30	42	16.8

Sub-test No.V. PROBLEMS

1	197	78.8
2	122	48.8
3	160	64.0
4	92	36.8
5	182	72.8

(Contd.)

1	2	3
6	167	67.8
7	62	24.8
8	112	44.8
9	172	68.8
10	100	40.0
11	110	44.0
12	160	64.0
13	85	34.0
14	147	58.8
15	132	52.8
16	155	62.0
17	115	46.0

Sub-test Number VI. NUMBER SERIES

1	137	54.8
2	135	54.0
3	60	24.0
4	85	34.0
5	55	22.0
6	90	36.0
7	67	26.8
8	80	32.0
9	65	26.0
10	35	14.0
11	35	14.0
12	27	10.8
13	55	22.0

(Contd.)

1	2	3
14	24	10.0
15	42	10.8
16	23	9.2
17	135	54.0
18	23	9.2
19	47	38.8
20	97	36.8
21	80	32.0
22	77	30.8
23	33	13.2
24	57	22.8

Sub-test No.VII. JUMBLED SENTENCES

1	165	66.0
2	130	52.0
3	137	54.8
4	140	56.0
5	135	54.0
6	140	56.0
7	155	62.0
8	157	62.8
9	130	52.0
10	150	60.0
11	160	64.0
12	127	50.8
13	135	54.0
14	130	52.0

(Contd.)

1	2	3
15	140	56.0
16	140	56.0
17	125	50.0
18	124	49.6
19	123	49.2
20	122	48.8
21	120	48.0
22	121	48.4
23	122	48.8
24	119	47.6

The items of the First Try Out Testing containing sub-tests administered to the pupils of the schools mentioned and their actual responses were recorded so that inference could be made there unto for further try outs.

Here is the actual record of the actual work done over the seven sub-tests.

Analysis of the Data

The item modification.- The responses were scored as right or wrong.

1.Irrespective of age, frequencies of correct responses, wrong responses and items not attempted, responses for each item were counted.

2. The sub-test consisting of Problems posed
difficulties to children.

The pupils found difficulties in problems posed
in the sub-test because they were not accustomed to such
tests. Moreover their various practices followed in the
day-to-day work of the school, their way of completing
home work or assignment work had inhibitions . They could
not pick up a few items because the form of the items was
a bit strange to them. All and sundry reasons led the
investigator to change the form and arrangement of the
items for the First Try Out and subsequent ones.

The following table gives the number of detailed
items and the composition of the tests for the Second Try Out.

T A B L E 5

Table showing the Number of Items in each sub-test in the
First Try Out with Number of items retained

Sr. No.	Name of Sub-tests	Original items	Deleted Items	Percentage of Rejection	Retained for next Try Out
1.	Opposites	40	12	30.0	28
2.	Similarities	40	11	27.5	29
3.	Classification	30	3	10.0	27
4.	Analogies	40	10	25.0	30
5.	Problems	30	13	43.3	17
6.	Series	27	3	11.1	24
7.	Jumbled sentences	30	6	20.1	24
	Total	237	58		179

Thus out of 237 items, 58 items were rejected
and 179 items were retained for the First Try Out Test.

Time Limit

Time limit is a very important point in the
Try Out, more important than the later use of the task
in its final form. The reason for this that the items
are arranged only in a rough order of difficulty in the
try out form. If the time interval is too short,
pupils may not have adequate time to try items at the
tail of the test. Perhaps, they could have answered
them correctly had they enough time.

As the chief aim in the try outs was to obtain
sufficient data for item analysis, a small group of
pupils was administered the test and then gradually
the number of pupils was raised in other trials in
second try out, Pilot Test and the final run. The
time limit was fixed on the basis of experience
gained in this.

About time limit, R.L. Thorndike says, "If the
test is primarily a power test time allotment is set
so that most of the individuals tested have opportunity
to attempt all or nearly all the test items",

On the same issue, C.C. Ross[1] recommends,
"Sufficient time to be allowed so that all or almost
all the pupils can finish".

---

[1]
Ross, C.C. : Measurement in Today's Schools,
New York; Prentice Hall Inc. p. 156.

TABLE 6

Average Time of the Sub-tests for First
Try Out Test

Sr. No.	Name of Sub-test	No. of Items	Average Mts.	Time Secs.
1.	Oppisites	40	15	- 5
2.	Similarities	40	18	- 43
3.	Classification	30	26	- 00
4.	Analogies	40	25	- 00
5.	Problems	30	20	- 12
6.	Number Series	27	10	- 00
7.	Jumbled Sentences	30	15	- 00
	Total	237	130	- 00

From the average time taken by the pupils to complete the sub-test, average time for the group was calculated. This was recorded after checking up the try outs and responses.

The group took two hours and ten minutes to complete the whole battery of tests.

In First Try Out of the Test on a small sample of 250 pupils covering the range of 12 plus to 16 plus was mainly used in studying the primary items. Out of 237 items, 179 items were finally selected for further try outs. The approximate time required for administering the total test was recorded.

R e f e r e n c e s

1. Guilford, J.P. :   Psychometric Methods, Ch.15,
                        New York; Mc Graw-Hill
                        Book Co. Inc.

2. Lindquist, E.F. :   Educational Measurement, Ch.8,
                        Washington, D.C.;
                        American Council on
                        Education

3. Ross, C.C. :   Measurement in Today's School,
                        New York; Prentice Hall,
                        Inc.

CHAPTER VI
THE SECOND TRY OUT OF THE TEST

Introduction

As mentioned in previous chapter, the first
try out was carried out in a few schools of Srinagar
to study how the test worked. To make the test more
efficient and valid, it was thought proper to administer
the second try out of the test in all districts of
Kashmir. Lindquist supports more try outs.
"More than two try outs may sometimes be deemed
advisable, particularly if the tests are to be used
to make important decisions about the examinees and
if the time and resources, that the final form of the
test will be highly efficient as well as valid"[1].

The Objectives of the Second Try Out Test

1. To identify weak and defective items.

2. To identify non-functioning or implausible
distractors in the multiple choice type test.

3. To provide data to determine the difficulty
level of individual items.

---

[1] Lindquist, E.F. : " Educational Measurement".
American Council of Education, Washington,
D.C. p.252.

4. To provide data to determine the internal consistency of the test.

5. To provide data to fix up the time limit for each sub test.

Administering the Test

Sample

As the results of the second try out would determine the quality and the nature of the items with respect to the population on which the norms are to be fixed, the sample used for try out testing was s-elected so as to resemble the sample of the population. To facilitate the calculation in the try out, the test is administered to a sample of 370 pupils, for, according to Guilford[2], " a sample of about 400 pupils would give reliable data for item analysis".

Kelley[3] has shown that the most accurate determination of item validation or internal consistency can be obtained by comparing approximately the upper and lower 27 per cent of the total group. It can be seen that 27 per cent of 370 is 99.9 or approximately 100.

Micheels[4] has expressed the same opinion and has suggested the same figure.

2.
Guilford, J.P. : "Psychometric Methods". McGraw-Hill Book Co. Inc., New York, p.293.
3.
Thorndike, R.L.: "Personnel Selection", John Wiley & Sons Inc., New York, 1949. p.345.
4
Micheels, W.J. and Karnes, M.R. : "Measuring Educational Achievements". McGraw-Hill Book Co. Inc., New York, 1950. p.461.

For the present sample, 370 pupils between the age group 12 plus to 16 plus were picked up from various schools of Kashmir.

Table given below shows the name of schools and the number of pupils where the test was administered.

T A B L E 7

S. No.	Name of the School	District	Age	Students
1.	Govt. High School Anantnag	Anantnag	12 plus	100
2.	Bigbhara High School	"	13 "	87
3.	Kulgam Girls High School	"	14 "	47
4.	Govt. High School Sopore	Baramulla	15 "	46
5.	National High School	Srinagar	16 "	90
	Total			370

Item Analysis

The main objective of this try out was to study the items for analysis.

To Cyril Burt goes the credit of establishing the fundamental principles of item analysis. In 1921, he carried out the analysis and revision of Binet-Simon Scale. After this, item analysis became the indispensable procedure in test construction. It serves many useful purposes in the techniques of test construction. The information obtained from the procedure of item analysis may be tested on the following:

1. It supplies information concerning the item as a whole.

2. It gives a measure of the correlation between the test item and the criterion.

3. It supplies measure of the internal consistency of the test.

4. It yields a measure of the difficulty of each item.

Useful surveys have been published on the construction of item analysis, thereupon by Long and others[5]. Davis[6], Guilford[7], Lindquist[8], Lawshe[9], Swinford[10], Zublin[11], and others. Vernon[12] has brought out all the important facts of item validity and item consistency. He advocates two methods for item validation.

A brief description of the methods is as under:

[5] Long and Others: The Validation of Test Items. Ontrice College of Education, Toronto, 1935.

[6] Davis, F.B.:Item Analysis Data - Their Computation, Interpretation and Use in the Test Construction" Howard Educational Paper 2, Cambridge,1945.

[7] Guilford, J.P. :Ibid.

[8] Lindquist, E.F.: Ibid.

[9] Lawshe, C.H. : A Monograph for Estimating the Validity of Test Items. Journal of Applied Psychology, 1942

[10] Swinford: Biserial Pearson's 'r' as a Measurement of Test Item Validity. Jr. of Edul. Pay.XXV,1936.

[11] Zublin, J. : The Method of Internal Consistency for Selecting Test Items. Jr. of Edul. Pay. XXV, 1934

[12] Vernon, P.E. : Indices of Item Consistency and Validity. Br. Journal Psychology Statictics Section, Part III, 1948, pp. 152-167.

(a) Grouping Methods

In these methods, the criterion scores are divided into two or more categories. The right and the wrong responses for each item are recorded separately from the answer sheets of each subject. The whole sample can, thus be split up into two or more groups as follows:

1. The highest and the lowest 27 per cent.

2. The number of the student above and below the median.

3. The highest and the lowest quartiles or 10 per cent.

4. Three groups - high, medium and lower, as suggested by Vernon.

The main Groupting Methods are listed below:

1. Chi-square Method . - Chi-square or analysis of variance can be used for testing the significance of the difference in 'P' between successive groups. It can be applied when upper and lower 27 per cent groups are to be compared.

2. Graphical Method.- In this method, four more graphs are drawn for each separate item of the percentage of the passes. The unsatisfactory items can easily be judged from the graphs.

3. Percent Difference Method . - In this method, indices which amount to the difference of the percentage of passing of the high and low groups are suggested by Symonds & others.

4. The Constant Method. - Ferguson and Lawley derived the formula for this technique. This is a psycho-physical method.

5. Tetra-Choric Correlation . - Vernon has determined two values of correlation by contrasting the percentage of passing in the top tertile and two bottom tertiles, percentage of passing in the top tertiles with the percentage of bottom tertile. Average of the two coefficient yields a single discriminative index which has been called by Vernon, "The double Tetra-Choric". This difference between the coefficients will indicate whether the item is more effective in the upper part or in the lower part of the scale

6. Sigma Difference Method . - Lawshe and Mayer[13] employ a monograph to read off the value of the difference of the percentage of passing of the highest and the lowest groups in terms of standard deviation.

7. Biserial 'r' Method . - Flagman has prepared the biserial 'r' of the difference of percentage of passing the highest and the lowest groups.

8. Z Method . - Davis[14] has prepared the table of to road the Fisher's Z value for corresponding biserials.

13
Lawshe, C.H. and Mayer, J.S. : Students in Item Analysis of Item Validation on Test Reliability - The Effect of Two. Journal of Psychology, XXXI, 1947. pp. 271-77

14
Davis, F.: Item Analysis Data - Their Computation, Interpretation and Use in Test Construction. Harvard Education Paper No 2 Cambridge 1947

Distribution Methods

1.   The Standard Deviation Between the Means. -
This method is suggested by Burt as an alternative
method of expressing the "percentage overlapping". It
is used where an item is assumed to yield two point
distributions.

2.   Overlapping Method. - Guilford, Long and
others have developed similar methods from the proport-
ion of the wrong whose criterion score exceeds the median
score of the right. This method is out of use in
these days.

3.   Anstey Method. - The simple difference between
the mean criterion score of the pass and fail item was
first suggested by Swinford. It is the chief method
used by Anstey for item analysis. It is also called
Anstey D-Method of Item Analysis.

4.   Biserial Correlation Method. - This is the
most popular method which is frequently used by the
test constructors, but it has a main disadvantage that
it gives a high correlation.

The Two Methods Compared

Distribution Methods are more useful when the
number of the subjects is small and when the detailed
analysis of the response is required. The same method
can be used ·when the time limits have been marked
insufficient. Anstey D-Method is the simplest method

for the high and low difficult items, but the biserial
'r' gives indices independent of the balance.

The Grouping Methods are much quicker. Out of
all the grouping methods Tetra-Choric Method requires
the smallest number of cases.

In general, Burt points out, "There can be no one
best procedure. What is best in any particular case
can be decided by careful consideration of (i) the
special aim in view, and (ii) the type of data available.

Methods Used in the Present Work

It is obvious from the study of the various methods
that they are more or less suitable in any specific
situation. The sample being small, grouping upon by
dichotonising the groups will be least suitable.
Considering the above views of Distribution Methods
and Grouping Methods, it was decided to use the
following method :

> Biserial 'r' Method by using the tables
> prepared by Flanagenen.

> Details of Biserial 'r' Method as applied
> in the present task.

In this method, the discriminating value of each
item is calculated by finding the item - total test
correlation. An item yielding high biserial 'r'with
the total test is more discriminating than one that
yields low biserial 'r'. The criterion selected in
this method is the total score of the test.

The answer sheets of 370 pupils were arranged in an
ascending order according to the total scores secured.
The top 27 per cent of 370 is 100, i.e. 100 sheets from
above were taken and an equal number from below were taken
apart. The upper 27 per cent constitute the group of
pupils with high scores and the lower 27 per cent consti-
tute the group of pupils with low scores on the basis of
total scores on the test. Next the p e r c e n t a g e of
pupils in each group giving correct answers to an item was
calculated. As each group contains 100 pupils, the
number of pupils giving correct answers to each items gave
the percentage itself. These percentages were corrected
for guessing by the formula [15]

$$P_u = \dfrac{Ru - Wu}{\dfrac{K \quad - 1}{N \quad -NR}} \times 100$$

Where P_u = the percentage of correct responses in the
upper group after correlation for guessing.

Ru = the number of pupils in the upper group who
mark the item correctly.

Wu = the number of pupils in the upper group who
mark the item incorrectly

N = number of pupils in the upper group

NR = number of pupils who do not complete the item
in time.

In the present case, $NR = 0$ as liberal time was
given. A similar formula was used to apply correction

15

Lindquist, E.F. : " Educational Measurement.
American Council on Education, Washington, D.C.p.284.

for guessing to the percentages in the lower group also.

Having calculated the percentages, the indices of internal consistency or item discrimination were found out using the table prepared by J.C. Flanagan[16]. In this table are given the indices of internal consistency for each item. These indices in fact represent the item total test correlations for the different items. It can be seen from the table that all the items show a positive value of 'r'. This is natural as the validity of all the items have already been tested against the external criterion and items lacking validity have already been screened out as shown below. The high positive value of 'r' for almost all the items show that all the items are to a more or less extent homogeneous. According to Thorndike[17] "a correlation coefficient of .25 represents an outstanding validity". In the present case, thirtyone items are rejected as they show internal consistency values of less than .25.

Difficulty Levels of the Items

Different investigators have recommended different methods to compute the item difficulty indices.A

---

[16]
Thorndike, R.L. : Personnel Selection". John Wiley and Sons Inc., New York.p. 348.

[17]
Ibid., p. 245.

common method is to calculate the percentage of students answering an item correctly, and to judge the difficulty from the percentage of testees giving the correct answer to the item. A second method which has become very popular with test constructors is to use the extreme scores of the distribution.

Generally the upper and the lower 27 per cent of the distribution are used to compute item difficulty. In the present test, this method is based to compute the difficulty value of the item. The formula used to find the Difficulty Value is -

$$D = \frac{U + L}{2}$$

where D = Difficulty Value

U = Percentage of students scoring the item correctly in the upper group

L = Percentage of students scoring the item correctly in the lower group.

Doubts About Difficulty Values

Some doubts are expressed about the reliability of difficulty values computed by this method as it involves the elimination of the middle 46 per cent testees. F. Davis[18] has investigated the problem and has concluded that, "the less of reliability incurred by estimating indices from only 54 per cent of the sample is not sufficient to be of practical consequence

---

[18] Lindquist, E.F. : Op.cit., P.283

when the two criterion groups employed include at least 100 examinees".

The same investigator further says,"experimental evidence has shown that the difficulty indices of the sort are extremely reliable when they are based on samples as large as 400[19]: The sample, in the present case, consists of 370 testees. The reliability of difficulty indices calculated by this method can therefore be vouchsafed. The internal consistency indices and the difficulty values of the items are given in the table. The lower the value of D, the higher the difficulty level.

Methods of Scoring

The major point in objective type of questionning is to decide the method of scoring. There are certain types of tests where pupils can answer simply by guessing. Such forms in the present test are multiple choice type and two alternative type. The point to be decided is whether a raw scores obtained on these types should be corrected for guessing or not. Dr. Micheels and W. Karnes[20] giving their views on this question, write, "There does not seem to be much evidence to indicate that the type of test is significantly improved by correcting for guessing. The authors of the same are of the opinion that little is to be gained in

19
   Ibid., p. 263
20
   Micheels, W.J. and Karnes: Measuring Educational Achievement. McGraw-Hill Book Co., New York, 1950.

using correlation formula. But for item analysis,
Lindquist[21] suggests that "correlation for chance
success should be made use of individual raw scores
that are to be used for internal consistency purposes".

In the present test no correlation for chance
success is applied while scoring only for internal
consistency purposes, following Lindquist correction
is applied.

---

[21]
    Lindquist, E.F. : A First Course in Statistics.
Houghton Miffin, Boston, 1942.

TABLE 8

Internal Consistency Data of the
Difficulty Values of Items in
Seven Sub-tests

Item No.	u	L	'r'	D	Remarks
	Sub-Tst No. I. OPPOSITES				
1	78	51	.30	54.5	
2	32	24	.10	28.0	Rejected
3	27	9	.29	18.0	
4	62	26	.37	44.0	
5	79	42	.39	60.5	
6	56	30	.27	43.0	
7	60	42	.13	51.0	Rejected
8	57	26	.32	41.5	
9	59	28	.32	43.5	
10	80	46	.37	63.0	
11	28	18	.13	23.0	Rejected
12	49	30	.20	37.5	,,
13	58	30	.29	44.0	
14	80	47	.37	63.0	
15	66	41	.26	53.5	
16	72	48	.26	60.0	
17	33	24	.11	28.5	Rejected
18	78	52	.29	65.0	
19	75	22	.53	48.5	
20	61	43	.13	52.0	Rejected
21	54	28	.27	41.0	

1	2	3	4	5	6
22	64	20	.46	42.0	
23	36	25	.13	30.5	Rejected
.24	74	50	.26	62.0	
25	54	30	.25	42.0	
26	54	30	.25	42.0	
27	38	9	.40	23.5	
28	34	6	.44	20.0	

Sub-test No. II. SIMILARITIES

1	2	3	4	5	6
1	61	30	.32	45.5	
2	78	52	.29	65.0	
3	64	22	.43	43.0	
4	70	32	.38	51.0	
5	62	30	.33	46.0	
6	52	28	.26	40.0	
7	33	24	.11	28.5	Rejected
8	80	46	.37	63.0	
9	75	22	.53	48.5	
10	63	30	.34	46.5	
11	54	28	.27	41.0	
12	64	20	.46	42.0	
13	70	35	.35	52.5	
14	32	24	.10	28.0	Rejected
15	62	26	.37	44.0	
16	38	8	.42	23.0	
17	32	8	.37	20.0	
18	58	24	.25	46.0	
19	69	30	.39	47.5	

1	2	3	4	5	6
20	64	60	.04	62.0	Rejected
21	60	33	.28	46.5	
22	71	27	.44	49.0	
23	81	29	.52	55.0	
24	73	49	.25	56.0	
25	54	28	.27	41.0	
26	84	17	.64	51.9	
27	43	29	.15	36.0	Rejected
28	79	52	.30	65.5	
29	80	26	.54	53.0	

Sub-test No. III   CLASSIFICATIONS

1	2	3	4	5	6
1	78	52	.29	65.0	
2	22	6	.31	14.0	
3	60	30	.31	45.0	
4	73	50	.25	61.5	
5	25	8	.29	16.5	
6	19	4	.37	11.5	
7	83	40	.46	61.5	
8	31	6	.43	18.5	
9	70	30	.40	50.0	
10	56	32	.25	44.0	
11	33	24	.11	28.5	Rejected
12	80	46	.37	63.0	
13	78	51	.30	64.5	
14	39	32	.08	35.5	Rejected
15	77	57	.25	67.0	
16	70	28	.42	49.0	

(Contd.)

1	2	3	4	5	6
17	60	22	.40	41.0	
18	41	12	.37	26.5	
19	30	12	.26	21.0	
20	34	14	.27	24.0	
21	30	9	.32	19.5	
22	63	22	.42	42.5	
23	33	10	.33	21.5	
24	38	9	.40	23.5	
25	54	30	.25	42.0	
26	78	50	.31	61.0	
27	34	6	.44	20.0	

Sub-test No. IV. ANALOGIES

1	2	3	4	5	6
1	70	32	.38	51.0	
2	72	40	.33	56.0	
3	90	60	.40	75.0	
4	72	48	.26	60.0	
5	74	50	.26	62.0	
6	60	20	.40	40.0	
7	60	24	.37	42.0	
8	76	52	.25	64.0	
9	32	8	.37	20.0	
10	69	30	.39	47.5	
11	64	20	.46	42.0	
12	32	24	.10	28.0	Rejected
13	27	9	.29	18.0	

(Contd.)

1	2	3	4	5	6
14	62	26	.37	44.0	
15	38	8	.42	23.0	
16	38	20	.22	29.0 Rejected	
17	34	14	.27	24.0	
18	46	12	.41	29.0	
19	80	52	.31	66.0	
20	28	4	.44	16.0	
21	48	37	.12	42.5 Rejected	
22	38	9	.40	23.5	
23	33	24	.11	28.5 Rejected	
24	72	12	.61	42.0	
25	70	35	.35	52.5	
26	27	9	.29	18.0	
27	42	18	.30	30.0	
28	43	29	.15	36.0 Rejected	
29	54	30	.25	42.0	
30	38	9	.40	23.5	

Sub-test No. V. PROBLEMS

1	54	28	.27	41.0	
2	64	20	.46	42.0	
3	70	35	.38	52.5	
4	78	51	.30	64.5	
5	32	24	.10	28.0 Rejected	

(Contd.)

1	2	3	4	5	6
6	62	26	.37	44.0	
7	42	18	.30	30.0	
8	36	25	.13	30.5	Rejected
9	49	22	.30	35.5	
10	41	12	.37	26.5	
11	48	37	.12	42.5	Rejected
12	38	9	.40	23.5	
13	54	30	.25	42.0	
14	43	27	.15	36.0	Rejected
15	78	50	.31	61.0	
16	34	6	.44	20.0	
17	82	42	.43	62.0	

Sub-test No.VI. ARITHMETICAL SERIES

1	2	3	4	5	6
1	78	52	.29	65.0	
2	22	6	.31	14.0	
3	60	30	.31	45.0	
4	73	50	.25	61.5	
5	25	8	.29	16.5	
6	54	28	.27	41.0	
7	64	20	.46	42.0	
8	72	12	.61	42.0	
9	70	35	.35	52.5	
10	27	9	.29	18.0	
11	32	24	.10	28.0	Rejected
12	48	37	.12	42.5	,,
13	68	22	.47	45.0	

(Contd.)

1	2	3	4	5	6
14	70	28	.42	49.0	
15	30	12	.26	31.0	
16	46	12	.41	29.0	
17	80	52	.31	66.0	
18	28	4	.44	16.0	
19	82	30	.53	56.0	
20	49	30	.20	37.5	Rejected
21	61	30	.32	45.5	
22	79	42	.39	60.5	
23	60	42	.13	51.0	Rejected
24	59	28	.32	43.5	

Sub-test No.VII. JUMBLED SENTENCES

1	2	3	4	5	6
1	30	9	.32	19.5	
2	61	30	.32	45.5	
3	79	42	.39	60.5	
4	56	30	.27	43.0	
5	60	42	.13	51.0	Rejected
6	57	28	.32	41.5	
7	59	28	.32	41.5	
8	72	48	.26	60.0	
9	58	30	.29	44.0	
10	28	18	.13	23.0	Rejected
11	63	22	.42	42.5	
12	33	10	.33	21.5	
13	48	37	.12	42.5	Rejected
14	38	9	.40	23.5	

(Contd.)

1	2	3	4	5	6
15	54	30	.25	42.0	
16	78	50	.31	61.0	
17	34	6	.44	20.0	
18	24	8	.28	16.0	
19	43	29	.15	36.0	Rejected
20	82	42	.43	62.0	
21	84	33	.32	48.5	
22	66	18	.49	42.0	
23	80	52	.31	66.0	
24	28	4	.44	16.0	

Study of the Table

The table seemed to be useful in ascertaining whether a particular item should be retained or should be rejected. In the present work items with coeÉficient or correlation less than 0.25 were rejected.

In each sub-test 4-5 items are indicated less than acceptable index of 0.25. It was decided to discard them in the final form of the test.

In all 31 items were rejected. It can be seen from the table of sub-test V that in the problems the coefficient of correlation of four out of seventeen items was less than 0.25. Thus four items were rejected and only thirteen items from that sub-test retained.

This sort of result led the author to
think about the reason. Thus the following can be
probable reasons:

Perhaps the reason for such rejection and
inability of the pupils in responding was that the
pupils were not accustomed to such tests. They
might have been used to solving the problems on
paper by mathematical calculations and not by oral
manipulations. That affected their speed of working
resulting in smaller output. Similarly the novelty
in the items of Analogies might have presented the
problem of phraseology in test items.

There were two options left to the author
of the test. Either to cancel the whole sub-test
or to add new items to the test. Second alternative
was to go on with the valid items with those sub-tests.
So it was decided to carry on with the valid items
and thus only valid items were retained.

The Pilot Test

Introduction

When a larger population of the testees is
to be covered up and their achievements are to be
recorded-it is better to have a pilot test. The
purpose of a Pilot Test is to make better ground
for the subsequent tests so that the difficulties that
would arise due to the environment of the population
concerned are minimised. After all this is the
project dealing with human material.

As is clear from the attempt in the second try out of 179 items 31 items were to be rejected on the basis of Internal Consistency and Difficulty Values. The experience gathered in the previous tests would go a long way in checking the Pilot Test. Then the Pilot Test was taken up in order that better administering of the test would be possible. In fact, it was found that though the number of items remained constant it was possible to administer the test to a larger group practically the same number of 370 students of each group, i.e. total of 1850 students.

In the first Try Out and Second Try Out there were 237 items and the test was administered to 250 students of different age groups. Out of 237 items 58 items were rejected and 179 items were retained on the basis of correct and incorrect responses from the students. Their percentage of the correct responses as well as the new order of their positions was established. While in the Second Try Out Test which was administered to 370 students of different age groups. Out of 179 items, 31 items were again rejected on the basis of internal consistency. Their time limit was also recorded.

The main object of Pilot Testing was to administer the same test comprising of 148 items to even larger group of 1850 students than the last one, i.e. 370 pupils of each age group from 12 plus to 16 plus on the following basis.

(1) To provide data to determine the approximate time limit, for the finished test.

(2) To determine the improvement for the efficient test administration.

(3) To determine whether the given directions are clear to the students and their improvement if needed.

Administering the Test

The test was administered to 370 pupils of each age group from 12 plus to 16 plus. This 1850 pupils were tested at this stage.

The following table shows the number of schools in the Pilot Test

TABLE 9

Number of Pupils Tested in Different Schools in

the Pilot Test

S.No:	Name of the School	District	Age	Students
1.	Nawakadal High School	Srinagar	12^+	370
2.	M.P.High School	-do-	13^+	370
3.	M.L.H.S.School	Anantnag	14^+	370
4.	High School Sopore	Baramulla	15^+	370
5.	S.P.H.S. School	Srinagar	16^+	370
	T o t a l			1850

The time for administering the test was fixed before hand by contacting the heads of the schools. This sort of psychological testing was a novelty for the students of Kashmir. The students

were told that this test was not for the examination
purpose and that they should not get nervous. This
assurance enabled them to attempt the test.

The test was conducted in a suitable manner.
The pupils were given the booklets containing various
sub-tests items and the various instructions pertaining
to the rules they were to observe while answering the
problems. As is clearly written in the booklets
they had to return those as soon as they finished
the test. There was no time limit fixed for the test.
They were told that they had to hand over the test
as soon as they complete the test. The author noted
the time taken by each pupil in answering the test.

This he did by nothing the time taken
individually by each pupil.

Time Limit for the Final Run

There are various views about fixing the
time limit of a test. Burt[22] suggests, "the time
allowed should be such that the individual will
almost but not quite finish the test within it".
Lawshe Opine[23] puts it, "As a general principle,
mental ability tests are the only one in which
speed is a factor. Consequently, the reader will
always consider speed in score determination.

---

[22] Burt, H.E. : "Applied Psychology". Prentice Hall
Inc., New York, 1953.p.421

[23] Lawshe, C.H. "Principles of Personnel Testing"
McGraw-Hill Book Co. Inc., New York.p.190.

However, some time limit is necessary. It is a customary to select such limits so that there is little emphasis on speed and that nearly every one can finish.

According to Ross,[24] "The time allowance for the test should be generous. Short time allowance should be avoided in order to secure the data needed for determining the difficulty and discriminating values of the items". Ruch suggests, "The time limit should be fixed so that 90 per cent pupil can attempt all items even within their powers.

Considering the above views, it was decided to allow liberal time limit.

The following table shows the number of students finishing the test in different time limits.

[24] Ross, C.C.: "Measurement in Today's Schools". Prentice Hall Inc., New York. p. 155.

TABLE 10

Frequency of Time Limit

Time in Minutes	No. of pupils completing the Test	Cumulative Frequency	Percentage of Cumulative Frequency
120	11	1850	100.0
110	22	1839	99.4
105	32	1817	98.22
100	30	1785	96.48
95	37	1755	94.88
90	36	1718	92.86
85	49	1682	90.92
80	122	1633	88.27
75	195	1511	81.67
70	168	1316	71.13
65	290	1148	62.06
60	284	888	46.38
55	269	574	31.02
50	151	305	16.49
45	106	154	8.32
40	22	48	2.59
35	26	20	1.4

Conclusions

It can be seen from the above table that -

1. the minimum time taken to complete the
test is 35 minutes and maximum time is 120 minutes.

2. 46 per cent pupils completed in 60 minutes,
80 per cent pupils completed in 75 minutes, and
94.4 per cent pupils completed in 116 minutes.

As there were 148 items selected for the
final run, with a view that about 90 per cent pupils may
complete the test in time, the time fixed for
the battery of tests for the final run was 100
minutes, i.e. one hour and forty minutes.

The items selected for the Pilot Test were
found to be suitable for the final test.

The tests were found to be objective in
character so that they could be applied in
varying conditions.

Booklets for the Final Run

About 6,000 test booklets were printed for
the final run. There are 148 items in all the
sub-tests.

References

1. Burt Cyril :   Hand Book of Test for Use in the
                       Schools,
                               Staples Press, London, 1948.

2. Cane, V.R. and   The Timing and Responses to
   Horn V.   :   Special Perception Questions,
                               Jr. of Expl. Psy.,1951, III.

3. Davis, F.B. :   Item Analysis Data - Their Compu-
                      tation, Interpretation & Use in
                      the Test Construction,
                               Harvard Educational Paper 2,
                               Cambridge, 1946.

4. Dessai, K.G. :   Construction and Standardization
                      of Group Tests of Intelligence,
                               Bharat Prakashan, Ahmedabad,
                               1954.

5. Garret, H.E. :   Statistics in Psychology and
                      Education,
                               Longmans Green and Company,
                               New York, 1953.

6. Guilford, J.P.:   Psychometric Methods,
                               McGraw-Hill Book Company,
                               New York, 1954.

7. Lawshe, C.A. :   A Nomograph for Estimating the
                      Validities of Test Items,
                               Jr. App. Psychology, 1942.

8. Lindquist, E.F.:   A First Course in Statistics,
                      Houghton Miffin, Boston, 1942.

9. Lindquist, E.F.:   Educational Measurements,
                               American Council on Education,
                               Washington, 1955.

10. Long and Others: The Validation of Test Items,
                      Ontric College of Education,
                      Toronto, 1935.

11. Mischeels, W.J.   Measuring Educational Achievement,
    and Karnes :   McGraw-Hill Book Company,
                               New York, 1950.

12. Swinford :   Biserial Pearson's 'r' as
                      Measures of Test Item Validity,
                               Jr. Edul. Psy, XXV, 1936.

13. Thorndike :   Personnel Selection,
                               John Viley and Sons Inc.,
                               New York, 1949.

14. Thorndike, R.L.   Measurement and Evaluation in
    and Hagen E.   :   Psychology and Education,
                                 John Wiley and Sons Inc.,
                                 New York.

15. Vernon, P.E.:   Indices of Item Consistency and
                                Validity,
                                 Br. Jr. of Psy. State,
                                 See Vol. I, Part III, 1948.

16. Zublin, J. :   The Method of Internal Consistency
                                for Selecting Test Items,
                                 Jr. Edul. Psy., XXV, 1934.