CHAPTER IV
PILOT TESTING OF IETAT

4.1   INTRODUCTION

After constructing a test, the next step is try-out of the test and this step is further divided into pilot testing and finalization of the test. While this chapter discusses the entire process of the pilot testing of the constructed IETAT, it also contains every detail about the finalization of the IETAT based on the data collected through the pilot testing.

4.2   PILOT TESTING

The items in the pre-pilot test form were arranged in an order which was thought to be proper. But then, this arrangement was an arbitrary one and so needed a revision so that the items would be in the order of difficulty i.e. the easier items would come first and the more difficult ones afterward. To find out the difficulty value of each test item, the test should be given to a small group representative of the population to be finally tested. Thus, the pilot testing of the test was done with the following objectives.

i)   To find the range of applicability of the test

ii)   To identify weak or defective items

iii)   To determine the difficulty value of each item included in the test.

iv)   To standardize the instructions to be given for the whole test.

v)   To find out if any item needs any change or modification in its structure.

vi)   To find out the discriminating power of each item so that the undesirables may be weeded out from the test.

vii)   To determine the validity of each item.

viii)   To fix the time limit for the whole test.

The complete process of pilot testing has been described below.

4.2.1   Sample for Pilot Testing

The IETAT was prepared for the pre-service teachers' population studying in B.Ed.-I [First Year] in the state of Gujarat. Thus, in terms to measure the aptitude of pre-service

teachers of Gujarat, all the pre-service teachers enrolled in B.Ed.-I during the academic year 2016-17 were considered as the population for the present study. For pilot testing purpose, the investigator had selected School of Education of Navrachana University, Vadodara, Gujarat, randomly by using lottery method. All the pre-service teachers studying in B.Ed.-I [First Year] were constituted as a sample for the pilot testing. The students were from various classes of society, habitat, level of study with varied family and educational background.

There were 50 pre-service teachers enrolled in the School of Education, Navrachana University, Vadodara, Gujarat, during the academic year 2016-17. Thus, all 50 pre-service teachers were considered as a sample for the preliminary testing. But during the time of pilot testing, 38 pre-service teachers were present in the institute while remaining 12 pre-service teachers were absent. Thus, the final sample size for pilot testing was restricted to 38 pre-service teachers and IETAT was administered to 38 pre-service teachers on 23rd January 2017. The test was administered in the morning session during 10:00 a.m. to 11:00 a.m. when the pre-service teachers being quite fresh to answer the test. The details regarding the characteristics of sampled pre-service teachers are given below:

4.2.1.1 Characteristics of Sample of Pilot Testing

The sample for the present study were pre-service teachers, important characteristics like age, gender, habitat, caste category, educational level, education stream and teaching experience which are considered for pilot testing have been depicted in the following figures 4.1 to 4.7.

Figure 4.1

Age-wise Distribution of Sample Selected for Pilot Testing
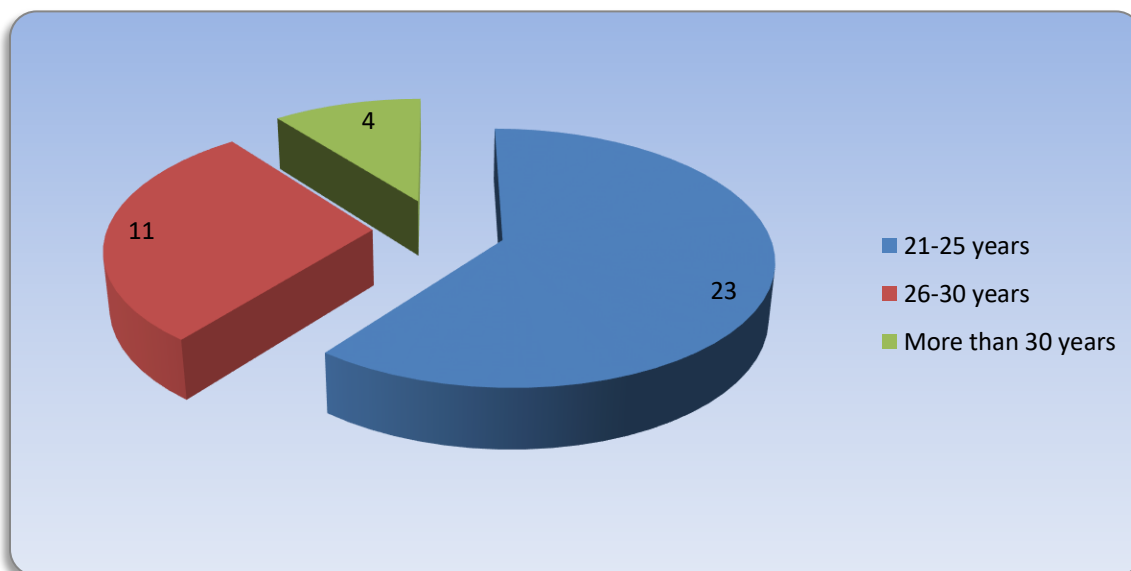
From the figure 4.1 it can be observed that out of 38 (100 percent) pre-service teachers selected for pilot testing of IETAT, 23 (60.53 percent) were of 21 to 25 years age group whereas remaining 11 (28.95 percent) and 4 (10.52 percent) were of 26 to 30 years and more than 30 years age group respectively.
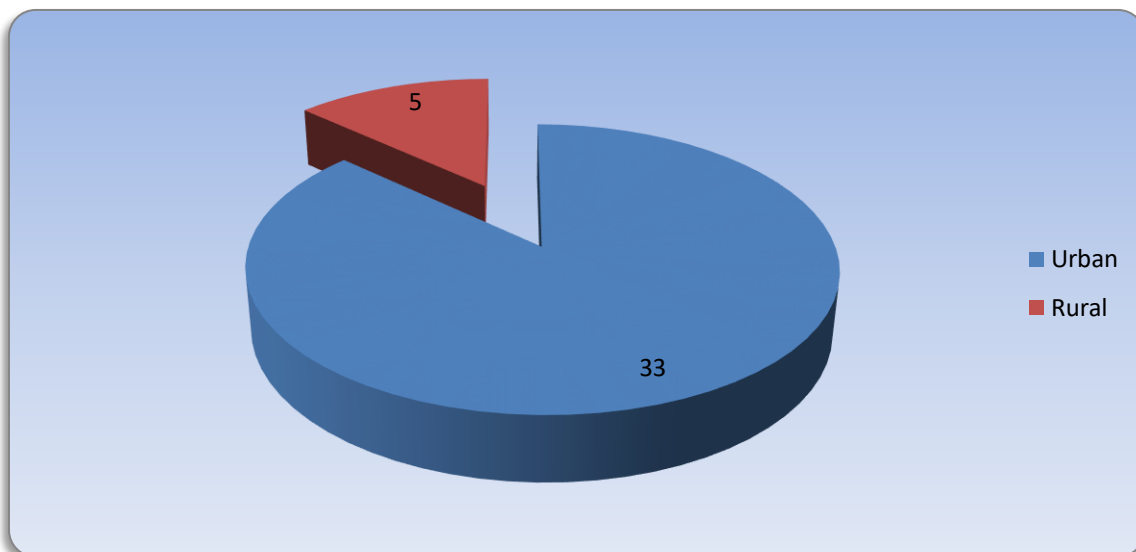
Figure 4.2

Gender-wise Distribution of Sample Selected for Pilot Testing
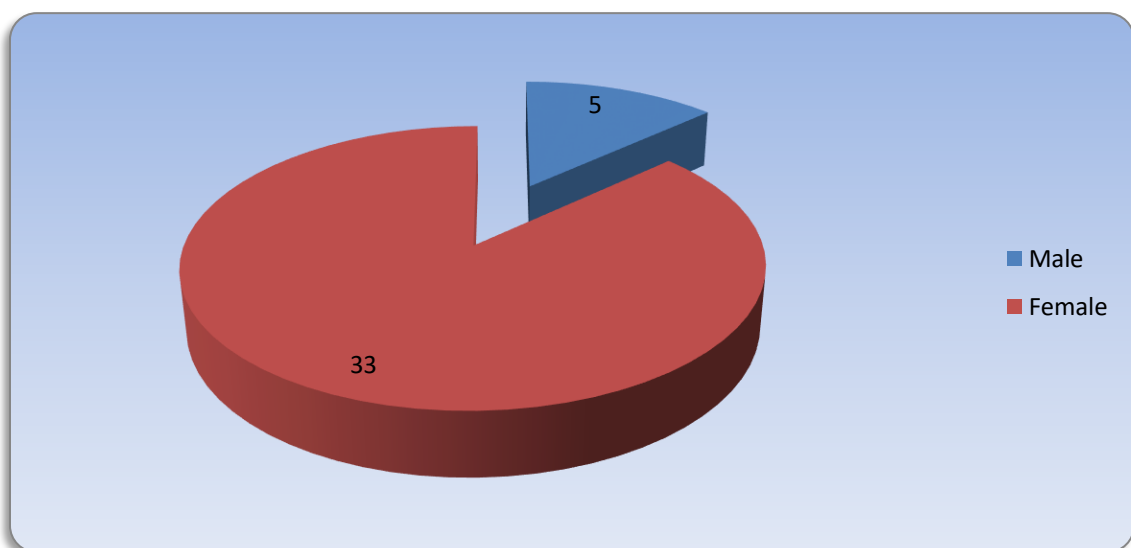
From the figure 4.2, it can be seen that out of 38 (100 percent) pre-service teachers selected for pilot testing of IETAT, most of them i.e. 33 (86.84 percent) were female whereas remaining 5 (13.16 percent) were male.

Figure 4.3

Habitat-wise Distribution of Sample Selected for Pilot Testing

From the figure 4.3, it can be observed that out of 38 (100 percent) pre-service teachers selected for pilot testing of IETAT, most of them i.e. 33 (86.84 percent) were from urban area whereas remaining 5 (13.16 percent) were from rural area.

Figure 4.4

Caste category-wise Distribution of Sample Selected for Pilot Testing

From the figure 4.4, it can be seen that out of 38 (100 percent) pre-service teachers selected for pilot testing of IETAT, most of them i.e. 37 (97.37 percent) were from general caste whereas only 1 (2.63 percent) was from other backward class.
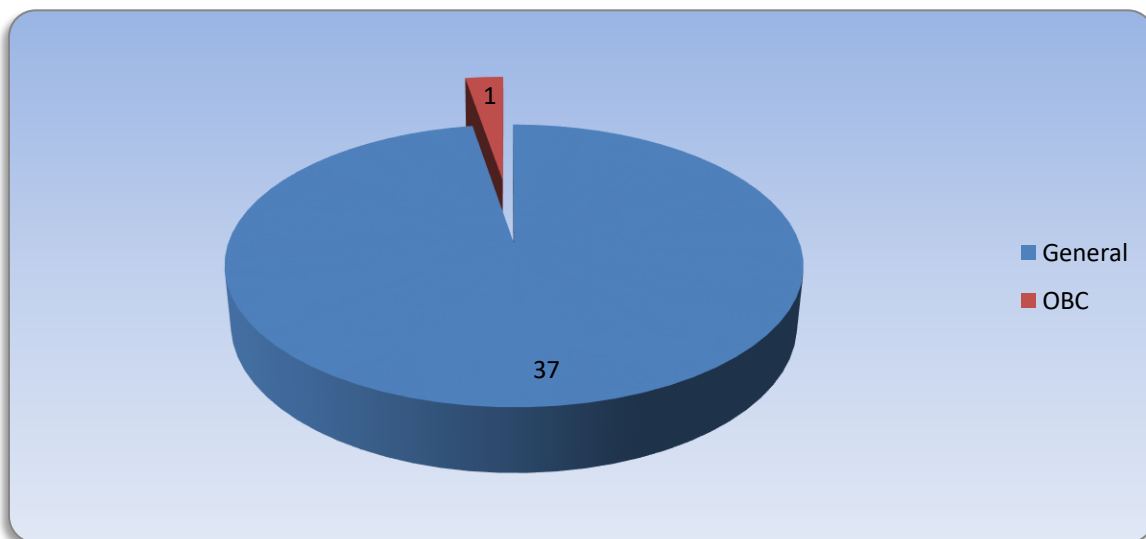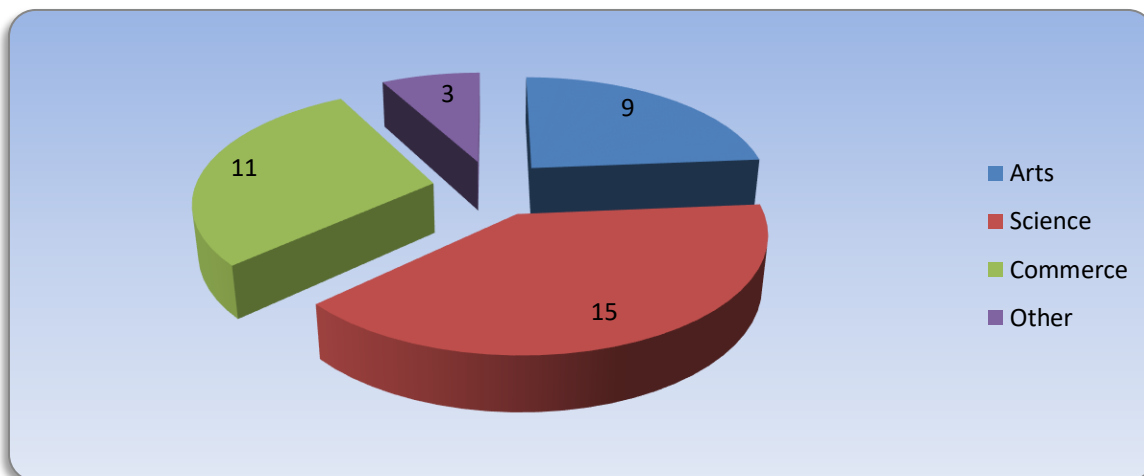
Figure 4.5

Educational Stream-wise Distribution of Sample Selected for Pilot Testing

From the figure 4.5, it can be observed that out of 38 (100 percent) pre-service teachers selected for pilot testing of IETAT, 15 (39.47 percent) were from Science stream whereas 11 (28.95 percent) and 9 (23.68 percent) were from Commerce and Arts stream respectively. The remaining and 3 (7.89 percent) were from other streams such as Engineering, Pharmacy and Business Administration.

Figure 4.6

Educational Level-wise Distribution of Sample Selected for Pilot Testing

From the figure 4.6, it can be seen that out of 38 (100 percent) pre-service teachers selected for pilot testing of IETAT, 24 (63.16 percent) had completed their graduation before joining B.Ed. whereas 14 (36.84 percent) had completed their post-graduation before joining B.Ed. Only 1 pre-service teacher had completed her M.Phil. before joining before joining B.Ed.
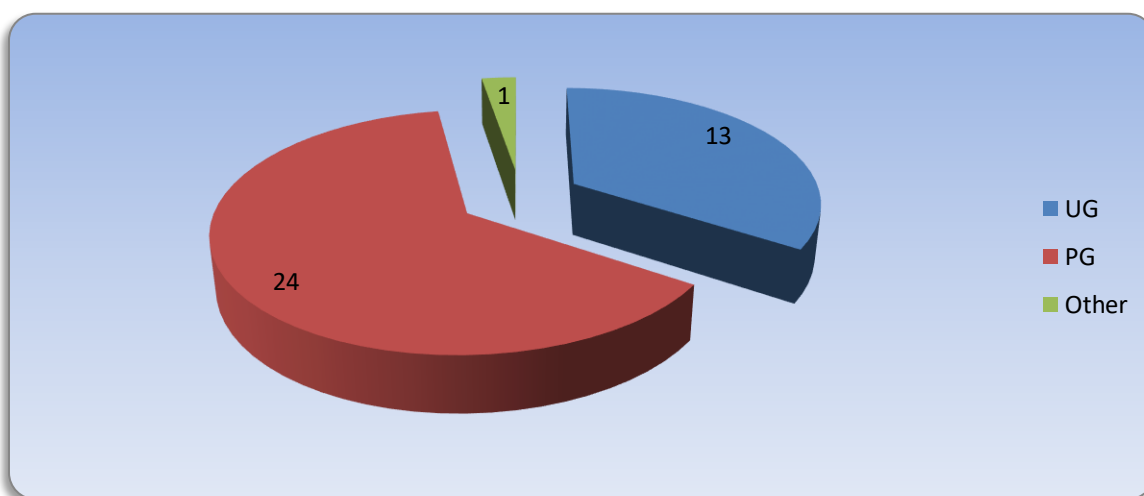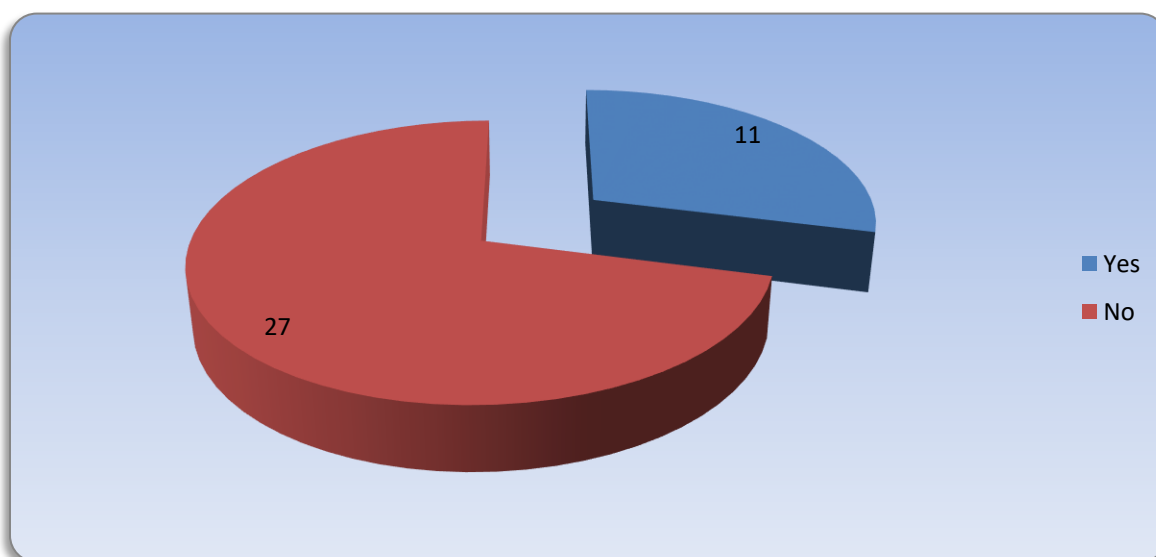
Figure 4.7

Teaching Experience-wise Distribution of Sample Selected for Pilot Testing

From the figure 4.7, it can be observed that out of 38 (100 percent) pre-service teachers selected for pilot testing of IETAT, a majority of them i.e. 27 (71.05 percent) were not having teaching experience prior to joining B.Ed. course whereas remaining 11 (28.95 percent) were having teaching experience prior to join B.Ed. course.

4.2.2 Instructions to Respondents

After review of the tests constructed earlier and consultation with the experts, the investigator prepared the final list of instructions for the respondents as follow:

❑ Provide all the general information about your age, gender, stream, level of education, teaching experience by tick (√) marking in the square box against the appropriate alternative given with every item.

❑ There are 50 items given under five factors viz. Knowledge about Inclusive Education (10 items), Perceived Ability to Identify Disabilities (10 items), Attitude towards Teaching Children with Special Needs (CwSN, 10 items), Perceived Ability to Adapt Inclusive Teaching Methods (10 items) and Skills to Manage Inclusive Classroom (10 items).

❑ Do not leave any item unanswered.

❑ There is no time limit for completion of this test. But work rapidly as much as you can

❑ The main purpose of this test is to measure your aptitude only. There are no marks for this test and this test will not affect your result or academics.

❑ Besides answering write difficulties (if any) in answering and your suggestions/opinions for further improvement of item(s).

❑ Research studies are useful only when reliable and accurate data are collected. So please give honest and sincere answers.

❑ Return the answer sheet along with the test-booklet to test administrator after answering all the items.

Though the written instructions are comprehensive and self-explanatory, the following oral instructions were also given.

❑ If you have any difficulty regarding the test then ask to test administrator but do not ask or discuss with others.

❑ You will be given enough time to answer all the items so give the answer after due thinking without hasting in finishing the test.

❑ You have to give answer of all the items so please see whether all items are answered before submitting your answer sheet to test administrator.

❑ You can write your suggestion(s)/comment(s) on item(s) for its further improvement.

❑ Your sincere and honest answers will help us a lot in our endeavor.

4.2.3 Administration of Pilot Test

The investigator had taken the permission from the Principal of Navrachna School of Education for pilot testing of IETAT (refer to APPENDIX VI). During the administration, the purpose of the test was made clear to all the 38 pre-service teachers. Necessary instructions were given to them orally before administration of the pilot form of IETAT (for detailed pilot test form, refer to APPENDIX VII). Time restriction was not implemented during the pilot testing of IETAT and the pre-service teachers were given chance to attempt every item of the test. but the time taken by the average number of students in attempting the whole test was noted down. The students were also requested to write the difficulties in answering the items and whole test. They were also asked to write their suggestions/comments on any of item or on the whole test in terms to improve the test further.

The test maker cannot control testing conditions but can take necessary precautions. Detailed instructions were provided in the test booklet and the sampled pre-service teachers had to follow the instructions strictly. The investigator had ascertained that the pre-service teachers understand the direction properly, don't use any unfair means and respond faithfully. Further, collecting data by the investigator himself secured the higher degree of uniformity in the administration of the test. Directions regarding answering items of test with example were given at the beginning of IETAT. Care was also taken to see that the directions provided in the test booklet are complete in respect of explaining to the pre-service teachers like what to be answered, how to answer and where to record the answers.

4.2.4 Time Limit

As it is exclusively a power test, time restriction was not imported and full time was given to the pre-service teachers to answer all the items of the test. As discussed in the previous chapter, the average time limit was fixed to 31.5 minutes (29 seconds per item) excluding the 3.5 minutes for reading instructions and 5 minutes for filling the general information. The average time limit, therefore, was tentatively fixed to be 40 minutes. The respondents were not informed about this time limit but they were instructed to raise

their hands as soon as they finish the whole test. The time was noted down when they started to answer. When the first hand raised the investigator noted the time and same as when the last hand raised. The shortest and longest time recorded in completing the whole test was 20 minutes and 40 minutes respectively.

4.2.5   Scoring

One mark was assigned to every correct answer of the item and no mark assigned to the wrong answer. The pilot form of test consisted a total of 70 items, thus, the total score obtainable was 70.

4.2.6   Correction of Chance Success

The pilot test consisted of multiple choice type 70 items and it was possible that some respondents may guess blindly among the choices and make the correct answer by chance. Test technicians argue that this type blind guessing can be reduced to some extent by warning respondents against it and allowing enough time to think the items critically. Following this, the respondents were advised not to answer by guessing and they were provided sufficient time to answer all the items. Therefore, no correction for chance success was applied while scoring. But following the suggestion of Davis (1959), correction of chance success was applied for internal consistency item analysis.

Davis (1959) suggested that it is especially important to make use of a correction for chance success in obtaining raw scores that are to be used for the internal consistency item analysis purpose. For this, the following formula II can be applied while scoring each respondent.

$$P_t = 100 \times \frac{R_t - \frac{W_t}{ki-1}}{N - NR_t} \qquad \text{(Formula II)}$$

Where

P_t = the percent of correct responses in the entire sample adjusted for chance success, and for omissions caused by not reaching the item in the time limit,

Rt = the number of examinees in the entire sample who *answer* the item correctly,

Wt = the number of examinees in the entire sample who answer the item incorrectly,

ki = the number of choices in the item,

Nt = the number of examinees in the entire sample,

NRt = the number of examinees in the entire sample who do hot reach the item within the time limit.

If every examinee has reached every time within the prescribed time limit, NRt becomes zero and computation of the adjusted percent is simplified accordingly. By applying the above formula II, the corrected percentages of the correct answers for each item have been given in table 4.1.

4.3   ITEM ANALYSIS

In order to produce effective and useful test, we must analyze the items from which the test is to be assembled. The items analysis is based on the statistical aspect of difficulty level and internal consistency indices. The main objective of item analysis is to obtain information concerning to items and thereby selecting the best items to compose the final form of test.

Aptitude tests are power tests so item analysis is more important than achievement type tests (Guilford, 1956) which are considered as speed tests. In this regard, Gulliksen (1950) says that in the construction of aptitude tests, the item statistics may be allowed to control rejection or selection of items more fully than in achievement tests.

As the present test is an aptitude test, it needs item analysis for composing the final test form.

After pilot testing, the scoring was done by the investigator himself. 38 answer sheets were examined and all the items were scored as per the scoring method described above. Based on the scores, two groups 'high scoring' and 'low scoring' were formed as follows:

❑   Ranks were assigned to answer sheets as per the score i.e. from highest to lowest.

❑ All the answer sheets were arranged as per the rank i.e. score sheet with highest rank at the top and with lowest rank at the bottom.

❑ From the pile of answer sheets, upper 27% (10 answer sheets with higher scores) and lower 27% (10 answer sheets with lower scores) were taken.

❑ The middle 46% (18 answer sheets) were discarded.

After the formation of upper and lower groups, the number of correct responses to an item in each group were found out and tabulated. From the correct response of each item, percentages were calculated. Then the correction of chance success was applied and percent of correct responses were calculated by using the formula II. The corrected percentages of correct responses have been given in table 4.1.

4.3.1 Item Difficulty

The standard method for determining the difficulty of items is the proportion of the group that answers the item correctly. When the item is scored either 0 or 1, the simplest index of its difficulty is the mean item score P (Guilford, 1956) and the most obvious way of expressing difficulty level of an item is the percent of the tryout group that marks it correctly (Davis, 1959).

Thus following these suggestions, indices of item difficulty for each item were calculated from the correct answer of upper and lower groups. A decrease in percentage will increase the difficulty value of an item.

The following formula III was used to calculate the difficulty value 'D' of each item.

$$D = \frac{U-L}{2} \qquad \text{(Formula III)}$$

Where,

D = difficulty value of the item

U = percentage of respondents scoring the item correctly in the upper 27% after being corrected for guessing work

L = percentage of respondents scoring the item correctly in the lower 27% after being corrected for guessing work

The difficulty values (D) for each item calculated by using the above formula are given in table 4.1. The lower the value of 'D', the higher the difficulty level of the item.

4.3.2 Item Discrimination Indices

This includes both 'internal consistency item discrimination' and 'item validity indices'. This discrimination may be in terms of total score on the test, or it may be in terms of some external criterion score of job performance. The relationships between the total score derived from a test and item scores are referred to as internal consistency item discrimination indices.

The present IETAT includes five sections to measure five different factors. So the test can be said as heterogeneous as it measures five different factors while the sections are homogeneous as the items include in each section are constructed to measure the same factor. Therefore, both, item validity and item analysis techniques were applied to the test items. Item validity was done by experts' judgment while item analysis was done by measuring internal consistency.

The commonly used methods to indicate the correlation of an item with the total score are bi-serial 'r', point bi-serial 'r', tetra-choric 'r' and the phi-coefficient. Out of these, the bi-serial correlation is usually regarded as the standard procedure in item analysis (Garrett, 1966) as an index of discriminating power appear to be most numerous. Thus, the bi-serial 'r' method was used to determine the discriminative power of the items of the present IETAT.

There are some formulas available for item analysis through bi-serial 'r' but it would be much laborious so the investigator used Flanagan's table of the normalized bi-serial coefficient of correlation as it makes simple to compute item validity coefficients from the percentages of correct answers in the upper and lower groups. The indices of internal consistency for each item are given in the following table 4.1.

Table 4.1

Internal Consistency Data [U & L], Internal Consistency Index [r] and Difficulty

Values [D] of the Items

Item No.	U%	L%	D	R	Item No. New Order
Section I: Knowledge about Inclusive Education					
1	**60**	**73.3**	**66.7**	**.00**	Rejected
2	86.7	50	68.4	.42	4
3	46.7	6.7	56.9	.53	6
4	60	6.7	33.4	.61	9
5	86.7	60	73.4	.34	2
6	73.3	50	61.7	.26	5
7	**100**	**100**	**100**	**.00**	Rejected
8	**33.3**	**33.3**	**33.3**	**.00**	Rejected
9	**-6.7**	**-6.7**	**-6.7**	**.00**	Rejected
10	100	46.7	73.4	.70	3
11	60	33.3	46.7	.29	7
12	100	73.3	86.7	.51	1
13	**33.3**	**20**	**26.7**	**.15**	Rejected
14	60	6.7	33.4	.61	8
15	33.3	6.7	20	.42	10
Section II: Ability to Identify Disabilities					
16	33.3	20	46.7	.53	15
17	33.3	6.7	20	.42	19
18	73.3	50	61.7	.25	12
19	**46.7**	**46.7**	**46.7**	**.00**	Rejected
20	73.3	46.7	61.7	.28	11
21	**-20**	**-6.7**	**-6.7**	**.00**	Rejected
22	**46.7**	**33.3**	**40**	**.10**	Rejected
23	73.3	20	46.7	.53	16

Item No.	U%	L%	D	r	Item No. New Order
24	20	6.7	13.3	.29	20
25	86.7	20	53.4	.62	14
26	33.3	6.7	20	.42	18
27	60	20	40	.40	17
28	**33.3**	**60**	**46.7**	**.00**	**Rejected**
29	86.7	33.3	60	.54	13
Section III: Attitude towards Teaching CwSN					
30	86.7	60	73.4	.32	23
31	60	20	40	.38	27
32	100	73.3	86.7	.50	21
33	100	20	60	.82	24
34	73.3	6.7	40	.71	26
35	20	6.7	13.4	.29	28
36	20	6.7	13.4	.29	29
37	20	6.7	13.4	.29	30
38	60	20	40	.38	25
39	**46.7**	**60**	**53.4**	**-.13**	**Rejected**
40	100	73.3	86.7	.54	22
41	**46.7**	**73.3**	**60**	**-.30**	**Rejected**
42	**73.3**	**86.7**	**80**	**-.18**	**Rejected**
Section IV: Ability to Adapt Inclusive Teaching Methods					
43	73.3	33.3	53.3	.41	33
44	86.7	60	73.4	.32	31
45	60	20	40	.38	35
46	**-6.7**	**6.7**	**00**	**.00**	**Rejected**
47	33.3	6.7	20	.42	39
48	46.7	20	33.4	.33	36
49	100	20	60	.80	32

Item No.	U%	L%	D	r	Item No. New Order
50	**6.7**	**-6.7**	**-6.7**	**.00**	**Rejected**
51	**6.7**	**20**	**-6.7**	**.00**	**Rejected**
52	**6.7**	**20**	**-6.7**	**.00**	**Rejected**
53	86.7	20	53.4	.60	34
54	**73.3**	**60**	**66.7**	**.15**	**Rejected**
55	46.7	6.7	26.7	.51	37
56	20	6.7	13.4	.29	40
57	46.7	6.7	26.7	.51	38
Section V: Skills to Manage Inclusive Classroom					
58	86.7	60	73.4	.31	42
59	**-6.7**	**-6.7**	**00**	**.00**	**Rejected**
60	**-20**	**-6.7**	**-6.7**	**.00**	**Rejected**
61	86.7	46.7	66.7	.45	43
62	73.3	6.7	40	.69	45
63	100	86.7	93.4	.38	41
64	33.3	6.7	20	.42	48
65	**6.7**	**6.7**	**6.7**	**.00**	**Rejected**
66	46.7	20	33.4	.29	46
67	33.3	6.7	20	.42	49
68	20	6.7	13.4	.29	50
69	73.3	46.7	60	.29	44
70	46.7	20	33.3	.29	47

From the table 4.1, it can be observed that out of 70 items covered under pilot testing, a total of 20 items were rejected due to low validity indices. Section wise, 5 (item 1, 7, 8, 9 and 13) from section I, 4 (item 19, 21, 22 and 28) from section II, 3 (item 39, 41 and 42) from section III, 5 (item 40, 50, 51, 52 and 54) from section IV and 3 (item 59, 60 and 65) from section V have been rejected due to their low validity indices.

From the same table 4.1, it can also be seen that the highest validity of an item was found to be .82 (item 33) and the lowest validity index was seen to be .26 (item 6). According to Garrett (1966), items with validity indices of .20 or more are regarded as satisfactory. Thus items with validity indices .20 or less than that were rejected.

Table 4.1 further shows the highest difficulty value of items retained, is found to be 93.4 (item 63) and the lowest difficulty value of the item seen to be 13.4 (items 35, 36, 37, 56 and 68). Thus the range of the difficulty values of the items selected was between 93.4 and 13.4 whereas the validity indices range was between .26 and .82.

All the retained items, then, rearranged as per their difficulty value i.e. from higher to lower difficulty value in each section. Thus the items were placed in order of most easy to most difficult.

4.4   ITEM SELECTION

The items for the present test were selected based on experts' criticism, item validity, difficulty value and internal consistency. All the items were validated against the item validity and internal consistency.

As described in the previous chapter, the items were referred to the experts for their criticism and based on their rating, CVR was found out for each item. 28 items were removed due to less content validity and 32 items were revised as per the suggestions of experts.

After calculating the difficulty values, the items were grouped as per the guidelines of Henning (1987) given in the following table 4.2.

Table 4.2

Distribution of Items of Pilot Test as per Henning's Guidelines of Item Difficulty

Difficulty Level	Description	Items	Total
≤ .33	High Difficult	4, 8, 9, 13, 14, 15, 17, 21, 24, 26, 35, 36, 37, 46, 47, 48, 50, 51, 52, 55, 56, 57, 59, 60, 64, 65, 66, 67, 68, 70	30
.34 to .66	Moderate Difficult	3, 6, 11, 16, 18, 19, 20, 22, 23, 25, 27, 28, 29, 31, 33, 34, 38, 39, 41, 43, 45, 49, 53, 62, 69	25
≥ .67	Low Difficulty/Easy	1, 2, 5, 7, 10, 12, 30, 32, 40, 42, 44, 54, 58, 61, 67	15
Total			**70**

Table 4.2 shows that out of 70 items, 15 were easy, 25 were moderate and 30 items were difficult items.

The data related to item difficulty value further grouped as per the scheme of distribution given by W. Summer and Garrett in terms to know the range of difficulty values of items.

Table 4.3

Distribution of Items of Pilot Test According to Difficulty Indices on the Lines of W. Summer

Difficulty Indices	Total No. of Items in Pilot Test		Total No. of Items Rejected		Total No. of Items Retained	
	No. of Items	% of Items	No. of Items	% of Items	No. of Items	% of Items
0 to 40	34	48.57	11	15.71	23	32.86
41 to 60	18	25.71	5	7.14	13	18.57
61 tp 100	18	25.71	4	5.71	14	20
Total	**70**	**100**	**20**	**28.56**	**50**	**71.43**

From the table 4.3 it can be seen that as per W. Summer's scheme of test item distribution, there should be 12, 37 and 13 items in the range of difficulty indices 0 to 40,

41 to 60 and 61 to 100 respectively. But the present test indicates that there were 34, 18 and 18 items in the range of difficulty indices 0 to 40, 41 to 60 and 61 to 100 respectively. Thus the distribution of items in the present pilot test was found somewhat different from the W. Summer's scheme of distribution.

The items were further grouped as per Garrett's scheme of test item distribution based on the difficulty indices. Garrett suggested the distribution of 25%, 50% and 25% of items in the range of difficulty indices 0 to 25, 26 to 75 and 76 to 100 respectively. The following table 4.6 presents the items of constructed IETAT according to the difficulty indices on the lines of Garrett's scheme of distribution of test items.

Table 4.4

Distribution of Items of Pilot Test According to Difficulty Indices on the Lines of Garrett

Difficulty Indices	Total No. of Items in Pilot Test		Total No. of Items Rejected		Total No. of Items Retained	
	No. of Items	% of Items	No. of Items	% of Items	No. of Items	% of Items
0 to 25	21	30	9	12.86	12	17.14
26 to 75	43	61.43	9	12.86	34	48.57
76 to 100	6	8.57	2	2.86	4	5.71
Total	**70**	**100**	**20**	**28.58**	**50**	**71.42**

From the table 4.4 it can be observed that as per Garrett's scheme of test item distribution, there should be 18, 35 and 17 items in the range of difficulty indices 0 to 25, 26 to 75 and 76 to 100 respectively. But the present test indicates that there were 21, 43 and 6 items in the range of difficulty indices 0 to 25, 26 to 75 and 76 to 100 respectively. Thus the distribution of items in the present pilot test is also not so close to Garrett's scheme of distribution.

Table 4.3 and 4.4 shows that the distribution items in the present test do not agree so closely either with the Summer's or Garrett's scheme of distribution. But it should be noted here that the reason of contrast lies in the selection or rejected of items. The items of the pilot test have been rejected or retained for the final test not on the basis of their difficulty indices but on the basis of the bi-serial coefficient of correlation values. Items having the 'r' at more than .20 have been selected for the final test while items having 'r' less than .20 have been rejected. Moreover, almost 55 items fall in the range between 20 and 93 'D'. This much range is sufficiently good and acceptable for any good predictor test.

The items were further categorized as per the guidelines of Ebel (1979) for discriminating power which is given in the following table 4.5.

Table 4.5

Distribution of Items of Pilot Test based on Ebel's Guidelines of Discriminating Power

Discriminating power	Items	Total	Remark
.40 and above	2, 3, 4, 10, 12, 14, 15, 17, 23, 25, 26, 27, 29, 32, 33, 34, 40, 43, 47, 49, 53, 55, 57, 61, 62, 64	28	Very good items
.30 to .39	5, 30, 32, 38, 44, 45, 48, 58, 63	9	Reasonably good items
.20 to .29	6, 11, 18, 20, 24, 35, 36, 37, 56, 66, 68, 69, 70	13	Need improvement
≤ .19	1, 7, 8, 9, 13, 19, 21, 22, 28, 39, 41, 42, 46, 50, 51, 52, 54, 59, 60, 65	20	Very poor items
Total		**70**	

Table 4.5 shows that out of 70 items, 28 items were very good whereas 9 items were reasonably good. 13 items were found that needed improvement whereas the remaining 20 items were found very poor which were completely eliminated.

Thus, the pilot testing of IETAT helped the investigator in receiving experience in administering the test, knowing the ambiguity and deficiency in some items, establishing rapport with the respondents and led to standardizing the instructions and time to be given while administering the pilot test.