Developing an Inclusive education teaching aptitude test: Pilot testing and item selection

Do not take the risk, pilot test first. -De Vaus (1993)

*Hemendra S. Mistry**

Abstract

Pilot testing refers to pre-testing of a particular research instrument. Pilot study is a crucial element of a good study design; it fulfils a range of important functions and can provide valuable insights for other researchers. This paper reports the pilot testing process of draft inclusive education teaching aptitude test (IETAT), principally focusing on the trial version of actual test administration along with item analysis in terms to determine difficulty values and internal consistency indexes of test items. Data were collected from 38 pre-service teachers of a teacher education institution. Result indicated that the pilot testing helped in removing the weak items with ambiguity and deficiency, standardisation of test instruction and time limit, and developing the final version of IETAT. This paper highlights the importance of pilot testing in terms of improving the test item validity, adds to the body of knowledge on pilot studies, and contributes to the development of test development.
Keywords: *Pilot testing, inclusive education, teaching, aptitude, pre-service teachers*

INTRODUCTION

After constructing a test, the next step is to try itout, and this step is further divided into a pilot study and the finalisation of the test. A pilot study is used in two different ways: feasibility studies, i.e., pilot-testing, or a trial run (Polit et al., 2001; Baker, 1994). The main advantage of pilot testing an instrument is to get advance warning about inappropriate or complicated testing. The major reasons for conducting a pilot study of a testing instrument can be testing adequacy, assessing feasibility, and designing a research protocol (Teijlingen Van & Hundley, 2001). Generally, pilot studies are likely to be "under-discussed, under-used, and under-reported" (Prescott & Soeken, 1989). Full reports of pilot studies are rare in the research literature (Lindquist, 1991; Muoio et al., 1995; Teijlingen Van et al., 2001). Research papers mostly refer only to one element, either pre-testing or pilot-testing, of an instrument for validity and reliability (De Vaus, 1993). When detailed pilot studies are mentioned in academic papers or reports, researchers get a chance to learn from the pilot study and get ideas for making necessary changes in their investigation (Teijlingen Van & Hundley, 2001). The process and outcomes of pilot testing described in detail can be very useful to others embarking on projects with similar methods and instruments. With the understanding that a well-designed and well-conducted pilot study can inform us about the best research process and, on occasion,

* DOYMI, Faculty of Education, University of Salamanca, Salamanca- 37008, Spain.

about likely outcomes, an attempt is made to report the entire phase of the pilot study, particularly the actual improvements made in the test.

Teacher education and school leadership are considered essential components by many academicians for implementing inclusive education in the classroom (Ainscow, 2005; Sandhill & Singh, 2005; Booth et al., 2003; Ainscow, 1991). The general teacher education diplomas and degree courses available in India were offering "special education" as an optional paper to train and prepare teachers to identify and assess disability, but it was not an integral part of the training and could not train teachers to deal with challenges, diversity, and negative attitudes (Singhal, 2005). This could have led to distrust in both the special and mainstream education systems, as well as keeping children with disabilities (CWD) at home for fear of abuse or neglect in the classroom (Zulka, 2005).If the teachers' attitude towards inclusion is not positive and they lack concern, then they find themselves unprepared for inclusion and for teaching all learners (Ellins & Porter, 2005; Forlin, 2001).

Until 2015, the teacher education programmes in India offered an optional paper on special education, and prospective teachers with an interest in inclusive education for CWD were opting for it. Since the National Council for Teacher Education's (NCTE, a statutory body of the Government of India) guidelines (2014), inclusive education has been an integral part of the curriculum. As pointed out earlier, since general teacher education courses are mainly focused on preparing teachers for general schools, there is a need to select candidates with aptitude towards inclusive education. Individuals with a high teaching aptitude for inclusive education should be identified through appropriate testing and advised to join an inclusive school after receiving training. As a result, admitting candidates with an aptitude for inclusion can help to ensure the success of inclusive education.

The general teacher education programmes (Diploma in Elementary Education [D.El. Ed.], Bachelor of Education [B.Ed.], and Master of Education [M.Ed.]) available in India are focused mainly on preparing teachers for general schools, and they merely teach inclusive education as a subject wherein pre-service teachers are equipped with theoretical knowledge but practical practise is not given its due (Sharma et al., 2009; Bhatnagar & Das, 2013). Thus, pre-service teachers often find themselves not trained enough for inclusive schools and thereby hesitate to join such schools (Forlin, 2001). This could be one of the main reasons for the shortage of teachers for inclusive education in India. Thus, selecting the right personnel for inclusive education to undertake teacher education courses through the application of suitable scientific techniques is the need of the hour.

When we say that a person possesses an aptitude for teaching in inclusive education, it is assumed that s/he has a good proportion of the traits that are required for becoming successful as an inclusive education teacher. The magnitude of these traits may differ from person to person, or even the number of traits possessed by each person may also differ, as some may possess more traits than others. A number of traits required for being a successful teacher in inclusive education compose, as a whole, the aptitude for teaching in inclusive education. Thus, the high or low aptitude for teaching in inclusive education is in proportion to the number of traits possessed by an individual. It also depends on the nature of the traits possessed.

When estimating the aptitude for teaching in inclusive education, the factors that contribute to the success of teaching in inclusive education should be measured through proper tests. The

aptitude for teaching inclusive education is in proportion to the number of such factors and also to their magnitude. Such factors are also important in conditioning success in teaching inclusive education. By constructing the present inclusive education teaching aptitude test, an attempt is made to satisfy a felt need for such a test. Unlike other tests constructed so far meant for general teaching aptitude, the Inclusive Education Teaching Aptitude Test (IETAT) was specifically prepared to measure the aptitude of pre-service teachers for teaching in inclusive education and is referred to as the IETAT.

The first phase of pilot testing of the IETAT involved discussions with experts to determine the factors related to successful teaching in inclusive education, the phrasing and order of items, and the range of answers on multiple-choice questions (Teijlingen Van & Hundley, 2001). A list of 33 traits relating to the teaching of inclusive education was prepared based on a review of literature and materials on teaching and learning in inclusive settings. The subject matter experts (SMEs) in the fields of inclusive education, teacher education, and psychology of education were requested to rate the most important traits for teaching in inclusive education. Besides rating, they were also asked to suggest traits that could be included in the IETAT. Based on the SME ratings, 12 traits were found to be most important. Furthermore, some of the interconnected and similar in nature traits were grouped, and the following factors were discovered:

- Knowledge about inclusive education
- perceived ability to identify disabilities
- attitude towards teaching children with disabilities
- perceived ability to adapt inclusive teaching methods
- Skills to manage an inclusive classroom

Initially, 97 items were framed (see author, 2018 for more information on the items) under the five factors listed above and referred back to the SMEs for criticism and content validity. Based on the SMEs' ratings of the items (Lawshe, 1975), 27 items with lower ratings were removed due to low content validity, and a total of 70 items were retained in the IETAT for pilot testing, with 32 items modified based on the SMEs' suggestions (see the author, 2018 for more details about the items).

Conducting a pilot study does not guarantee success in the main study, but it does increase the likelihood. Pilot studies fulfil a range of important functions and can provide valuable insights for other researchers. More discussion of the process and outcomes of pilot studies is needed among researchers (Teijlingen Van & Hundley, 2001). Thus, this paper attempts to discuss the process and outcomes of developing a valid test with the goals of determining the range of applicability, identifying weak or defective items, determining difficulty values and discriminating power of items, determining the validity of items, standardising instructions, and fixing the time limit for the entire test.

Methods

Participants

The sample for this pilot study was pre-service teachers enrolled in the first year of a two-year B.Ed. programme at a teacher education college of a private university, selected randomly by lottery. The college was not identified so as to maintain the anonymity of the respondents. There were 50 pre-service teachers enrolled in the teacher education college, and all were considered a sample for pilot testing. During the time of data collection, 38 pre-service teachers were present in the college, so the final sample size for pilot testing was restricted to 38 pre-service teachers. The demographic information of the sample is given in Table 1.

Table 1

Demographic information of sample

Variables	Demographic information
Age group	21 to 25 years (N=23, 61%) 26 to 30 years (N=11, 29%)
	More than 30 years (N=4, 10%)
Gender	Male (N=5, 13%) Female (N=33, 87%)
Habitat	Urban (N=33, 87%) Rural (N=5, 13%)
Educational level	Graduate (N=24, 63%) Postgraduate (N=14, 37%)
Education stream	Arts (N=9, 24%) Commerce (N=11, 29%) Science (N=15, 39%) Other (N=3, 8%)
Previous teaching experience	Yes (N=11, 29%) No (N=27, 71%)

Instrument

As no validated questionnaires relevant to the teaching aptitude of pre-service teachers towards teaching in inclusive education were found in the review of the literature, a draught of an IETAT was developed. The primary goal of the IETAT pilot testing was to determine item validity, item difficulty, a discriminating index for selecting items for the final version of the IETAT, standardise instructions, and determine testing time.70 items related to inclusive education were listed in the IETAT. Four options were given with every item. The respondents had to choose one correct option from the four options. The test also included items relating to the demographic information of the participants: gender, age, habitat (rural or urban), education level, educational stream, and previous teaching experience, if any. The items were organised under general information and five factors. The factor-wise distribution of the items is given in Table 2.

Table 2

Factor wise items included in the IETAT

	Factor	Item Nos. in IETAT	No. of Items
I.	Knowledge about inclusive education	1-15	15
II.	Perceived ability to identify disabilities	16-29	14
III.	Attitude towards teaching children with disabilities	30-42	13
IV.	Perceived ability to adapt inclusive teaching methods	43-57	15
V.	Skills to manage inclusive classroom	58-70	13
Total	**70 items**		

Procedure

A questionnaire survey design was employed to gather the data in terms of item difficulty and validity from the pre-service teachers enrolled in the first year of a two-year B.Ed. program. The detailed procedure of the pilot study is described below.

Administration of the Pilot Test

With prior permission from the administrative head of the selected teacher education institution, the pilot testing of IETAT was done in the second semester of the B.Ed. program. During the administration, the purpose of the test was made clear to the pre-service teachers. Necessary instructions (see under the heading "Instructions to Respondents" below) were given in the test and also verbally before administration of the pilot form of the IETAT. Time restrictions were not implemented during the pilot testing of the IETAT, and the pre-service teachers were given the chance to attempt every item of the test, but the time taken by the average number of respondents in attempting the whole test was noted down. Participants were also asked to write about their difficulties answering the questions and taking the entire test. They were also asked to write down any suggestions, comments, or feedback they had on any item or on the entire test in order to help it improve further. It is believed that the test maker cannot control testing conditions but can take necessary precautions. Detailed written instructions were provided in the test booklet, and the participant 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. Furthermore, having the investigator collect the data ensured a higher level of consistency in the test administration. At the start of IETAT, instructions for answering test items with examples were provided. Care was also taken to see that the directions provided in the test booklet are complete in terms of explaining to the pre-service teachers things like what to answer, how to answer, and where to record the answers. The test was administered to the pilot study participants in the same way as it will be administered in the main study (Peat et al., 2002).

Instructions to Respondents

After review of the tests constructed earlier and consultation with the experts, the following written instructions were given in the test:

- Provide all the general information about your age, gender, stream, level of education, and teaching experience by marking a "tick" (√) in the square box against the appropriate alternative given with every item.
- There are 70 items listed under five factors. You are required to respond to all questions.
- Do not leave any item unanswered.
- There is no time limit for completion of this test. However, work as quickly as possible.
- The main purpose of this test is to measure only your aptitude. There are no marks for this test, and this test will not affect your result or academics.
- Besides answering, describe your difficulties (if any) in answering and your suggestions, opinions, or feedback for further improvement of the 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 the test administrator after answering all the items.

Though the above-mentioned written instructions given in the test were comprehensive and self-explanatory, the following verbal instructions were also given to the participants.

- If you have any difficulty regarding the test, ask the test administrator; do not ask or discuss with others.

- You will be given enough time to answer all the items, so give the answer after due deliberation without hasting to finish the test.
- You have to give an answer for all the items, so please see whether all items are answered before submitting your answer sheet to the test administrator.
- You can write your suggestion(s) or comment(s) on the item(s) for its further improvement.
- Your sincere and honest answers will help us a lot in our endeavour.

Time Limit

As the IETAT was exclusively a power test, time restrictions were not imposed, and full time was given to the respondent to answer all the items of the test. The respondents were instructed to raise their hands as soon as they finished the whole test. The time was noted when they started to answer. The investigator noted the time when the first hand was raised, which was the same as when the last hand was raised. The shortest and longest times recorded for completing the whole test were 20 minutes and 40 minutes, respectively, and the average 30 minute time was reasonable (Peat et al., 2002).

Scoring

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

Results and Discussion

Item analysis

In order to produce an effective and useful test, the investigator analysed the items with which the IETAT is to be assembled. The item analysis is based on the statistical aspects of difficulty level and internal consistency indices. The main objective of item analysis is to obtain information concerning items and, thereby, select 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 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 according to rank, i.e., the score sheet, with the highest rank at the top and the lowest rank at the bottom.
- From the pile of answer sheets, upper 27% per cent(10 answer sheets with higher scores) and lower 27per cent% (10 answer sheets with lower scores) were chosen.
- The middle 46per cent% (18 answer sheets) were discarded.

The number of correct responses to an item in each group was determined and tabulated after the formation of upper and lower groups. From the correct responses for each item, percentages were calculated. Then the correction for chance success was applied, and the percent of correct responses were calculated using the above formula. The corrected percentages of correct responses have been given in Table 3.

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 the difficulty level of an item is the percent of the try-out group that marks it correctly (Davis, 1959). Thus, following these suggestions, indices of item difficulty for each item were calculated from the correct answers of the upper and lower groups. A decrease in percentage will increase the difficulty value of an item.

The following formula was used to calculate the difficulty value "D" of each item:

D= (U-L)/2

Where,

D = difficulty value of the item.

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

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

The difficulty values (D) for each item calculated by using the above formula are given in table Table 3. The lower the value of 'D', the higher the difficulty level of the item. The highest difficulty value observed was 93.4 (item 63), and the lowest difficulty value of the item seen was 13.4 (items 35, 36, 37, 56, and 68). Thus, the difficulty values of the items chosen ranged between 93.4 and 13.4, while the validity index ranged between 26 and 82. All the retained items were 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.

Item discrimination indices

This includes both "internal consistency item discrimination" and "item validity indices." This discrimination may be in terms of the 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 to be heterogeneous as it measures five different factors, while the sections are homogeneous as the items included 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 determined 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, which is usually regarded as the standard procedure in item analysis (Garrett, 1966), as an index of discriminating power, appears to be the most numerous. Thus, the bi-serial "r" method was used to determine the discriminative power of the items of the present IETAT.

The investigator used Flanagan's table of the normalised bi-serial coefficient of correlation, which makes it simple to compute item validity coefficients from percentages of correct answers in the upper and lower groups. The indices of internal consistency for each item are given in the following table:

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. Sectionally, five items (items 1, 7, 8, 9, and 13) from Section I, four (items 19, 21, 22, and 28) from section II, three (items 39, 41, and 42) from Section III, five (items 40, 50, 51, 52, and 54) from Section IV, and three (items 59, 60, and 65) from Section V were rejected due to low validity indices. Further, it can also be seen that the highest validity of an item was found to be 0.82 (item 33), and the lowest validity index was seen to be 0.26 (item 6). Garrett (1966) considers items with validity indices of.20 or higher to be satisfactory. Items with validity indices of.20 or lower were thus rejected.

Table 3

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: Perceived 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
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 children with disabilities					
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
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

Item selection

The items for the IETAT were selected based on experts' criticism, item validity, difficulty value, and internal consistency. All the items were validated against item validity and internal consistency. All unnecessary, difficult, or ambiguous items were discarded, and some items were reworded (Peat et al., 2002).

After calculating the difficulty values, the items were grouped as per the guidelines of Henning (1987) given in the following table 4. Out of 70 items, 15 were found easy, 25 were found moderate, and 30 items were found difficult. To determine the range of difficulty values of items, the data related to item difficulty values are further grouped according to the scheme of distribution provided by W. Summer and Garrett.

Table 4

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**

The data presented in Table 5 shows the distribution of items as per W. Summer's scheme. 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 to be somewhat different from W. Summer's scheme of distribution.

Table 5

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 to 100	18	25.71	4	5.71	14	20
Total	**70**	**100**	**20**	**28.56**	**50**	**71.43**

The items were further grouped as per Garrett's scheme of test item distribution based on the difficulty indices. Garrett suggested the distribution of 25per cent, 50per cent, and 25per centof items in the range of difficulty indices 0 to 25, 26 to 75, and 76 to 100, respectively. The following table 6 presents the items of the constructed IETAT according to the difficulty indices along the lines of Garrett's scheme of distribution of test items. 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 6

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**

Tables 5 and 6 show that the distribution items in the present test do not agree so closely with either Summer's or Garrett's scheme of distribution. But it should be noted here that the reason for contrast lies in the selection or rejection 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 their bi-serial coefficient of correlation values. Items with the "r" at more than 0.20 have been selected for the final test, while items with the "r" at less than 0.20 have been rejected. Moreover, almost 55 items fall in the range between 20 and 93 "D." This much variation is adequate and acceptable for any good predictor test.

The items were further categorised as per the guidelines of Ebel (1979) for discriminating power, which are given in Table 7. Out of 70 items, 28 were excellent, while 9 were adequate. 13 items were found that needed improvement, whereas the remaining 20 items were found to be very poor and were completely eliminated.

Table 7

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**		

Final IETAT

Based on the pilot testing, the final test has also continued to be in five sections. As against 15, 14, 13, 15, and 13 test items (making the total 70) of the five sections, respectively, of the pilot form, 10 items in each section (total 50) were retained in the final IETAT. The final version of the IETAT was translated into Gujarati and administered to 552 pre-service teachers in Gujarat, India. The reliability and validity of the IETAT were found to be 0.75 and 0.51, respectively. The test is intended to measure the teaching aptitude for inclusive education possessed by the pre-service teachers. In this way, the test anticipates the prospective teachers' future potentials and levels of capacity prior to their initial teacher training. The pre-service teachers with high aptitude identified through this IETAT will be most likely to acquire the skills and desired level of proficiency with a reasonable amount of training during their initial teacher training. Teachers with average and low aptitude levels should be provided more training on inclusion with practical exposure during their initial teacher training. This in turn will lead to preparing knowledgeable and skilled teachers for inclusive education and, to some extent, solve the problem of teacher shortage for inclusive education in India.

Table 8

Factor wise items in final IETAT

	Factor	Item Nos. in IETAT	No. of Items	% of Items
I.	Knowledge about inclusive education	1-10	10	20
II.	Perceived ability to identify disabilities	11-20	10	20
III.	Attitude towards teaching children with disabilities	21-30	10	20
IV.	Perceived ability to adapt inclusive teaching methods	31-40	10	20
V.	Skills to manage inclusive classroom	41-50	10	20
Total			**50 items**	**100**

Limitations

The main purpose of this pilot study was to assess the feasibility of IETAT and the validity of the items included in it. The pilot study of IETAT provided an opportunity to improve the test items. The pilot sample was confined to one teacher education institute only; hence, the data and findings were generated from one institution. This aspect may limit the generalizability of the findings, but looking at the similarity of the sample as possible to the target population, the result would not be very different from the large sample, subsequently reducing the threats to internal validity (Robichaud, 2016).

Conclusion

The pilot study undertaken to test the feasibility of the IETAT was vital for improving test validity and, subsequently, the usability of the test on a large sample. The planned procedures for test administration were also tested. The positive responses and relatively high response rate from pre-service teachers in the pilot confirmed the feasibility of administering the test to a larger sample using a standardised procedure, including time frame and instructions to be given. Thus, the pilot testing of IETAT helped the investigator receive experience in administering the test, knowing the ambiguity and deficiency

in some items, establishing rapport with the respondents, and standardising the instructions and time to be given while administering the final version of the test. This paper highlights the importance of pilot testing in terms of improving the test item validity, adds to the body of knowledge on pilot studies, and contributes to the development of test development.

References

Ainscow, M. 2005. From Special Education to Effective Schools for All. Key note presentation at the Inclusive & Supportive Education Congress 2005, University of Strathelyde, Glasgow.

Baker, T. L. 1994. Doing Social research (2nd ed.), McGraw-Hill Inc.New York.

Booth, T., Nes, K. &M. Stromstad. 2003. Developing Inclusive Teacher Education: Drawing the Book Together. In T. Booth, K. Nes, & M. Stromstad (Eds.), Developing Inclusive Teacher Education. (pp. 1-14). Routlege Falmer.

Davis, F. B. 1959. Item Selection Techniques. E. F. Lindquist (Ed.), Educational measurement, (pp. 266-328). American Council on Education.

DeVaus, D.A. 1993. Surveys in Social Research (3rd ed.), UCL Press.

Ebel, R. L. 1979. Essentials of Educational Measurement (3rd ed.). Prentice Hall.

Ellins, J. &J. Porter. 2005. Departmental differences in attitudes to special educational needs in the secondary school. British Journal of Special Education, 32(4), 188-195.https://doi.org/10.1111/j.1467-8578.2005.00396.x

Forlin, C. 2001. Inclusion: Identifying Potential Stressors for Regular Class Teachers. Journal of Educational Research, 43(3), 235-245.https://doi.org/10.1080/00131880110081017

Frankland, J. &M. Bloor. 1999. Some issues arising in the systematic analysis of focus group material. In R. Barbour &J. Kitzinger (Eds) Developing Focus Group Research.Politics, Theory & Practice, (pp. 144-155). SAGE

Garrett, H. E. 1966. Statistics in psychology and education (6th ed.). Longmans, Green and Co.

Guilford, J. P. (1956). Fundamental statistics in psychology and education.McGraw-Hill.

Gulliksem, H, 1950. Theory of Mental Tests. John Wiley & Sons.

Henning, G 1987. A Guide to Language Testing- Development, Evaluation, Research. Newbury House Publisher.

Lawshe, C. H. 1975. A Quantitative Approach to Content Validity. Personal Psychology, 28, 563-575. https://pdfs.semanticscholar.org/cd20/d9ae1554a2665983798a6f14a8f9a7272223.pdf

Lindquist, R. 1991. Don't forget the pilot work. Heart Lung: Journal of Critical Care, 20(1), 91-92.

Mistry, H. S. 2018. Construction and standardization of an inclusive education teaching aptitude test. An unpublished postdoctoral research report. The Maharaja Sayajirao University of Baroda. Vadodara.

Muoio, R., L. Wolcott, and H. Seigel. 1995. A win-win situation: The pilot program. Journal of Continuing Education in Nursing. 26(5), 230-233.

Peat, J., C. Mellis,K. Williams,&W. Xuan.2002.Health Science Research: A Handbook of Quantitative Methods. SAGE.

Polit, D.F., C. T.Beck, &B. P. Hungler.2001. Essentials of Nursing Research: Methods, Appraisal and Utilization(5th Ed.). Lippincott Williams & Wilkins.

Prescott, P.A. &K.L. Soeken. 1989. The potential uses of pilot work. Nursing Research, 38(1), 60-62. doi:10.1097/00006199-198901000-00015

Sandhill, A. &A. Singh.2005.Inclusion: Some Emerging Directions in the Indian Context. Paper presented at the Inclusive & Supportive Education congress 2005, University of Strathelyde, Glasgow.

Sharma, U., D.Moore, &S.Sonawane.2009. Attitudes and Concerns of Pre-service Teachers regarding Inclusion of Students with Disabilities into Regular Schools of Pune, Maharashtra. The Asia-Pacific Journal of Teacher Education, 37(3), 319-331. https://doi.org/10.1080/13598660903050328

Singhal, N. 2005a. Responding to Difference: Policies to Support 'Inclusive Education' in India.Paper presented at the Inclusive & Supportive Education congress 2005, University of Strathelyde, Glasgow.

Teijlingen Van, E. &V. Hundley.2001. The Importanceof Pilot Studies.Social Research Update, 35(35), 1-4. from https://www.researchgate.net/publication/11173521_The_Importance_of_Pilot_Studies