REFINEMENT OF MULTIPLE CHOICE ITEMS USING DIFFICULTY INDEX AND DISCRIMINATION POWER FOR ASSESSING MIDDLE STAGE LEARNERS' ATTAINMENT IN SCIENCE

Sunita Farkya[1], Indrani Bhaduri[2], Rachna Garg[3], Ashish Kumar Srivastava[4] and B. K. Tripathi*[5]

[1,3,4]Department of Education in Science and Mathematics
NCERT, New Delhi
[2]Educational Survey Division, NCERT, New Delhi-110016
[5]Director Inter University Centre for Teacher Education,
BHU, Varanasi

The present research study proposes to analyse multiple-choice items for the development of assessment tools for measuring conceptual learning in science. Initial pool of items were reviewed in a Focused Group Discussion (FGD) with a national level expert group and 44 items were eliminated out of 691 pool of items. Rest 647 items were again reviewed in a Focus Group Discussion (FGD) with state level experts for contextual and language intervention. As a result 30 items of total 647 items were modified. The finalised 647 items were subjected to field try out. The quantitative data were obtained by administration of total 647 items among 516 students from Grades 6, 7 and 8 theme-wise separately. Data obtained were subjected to individual item analysis in which both students' answers and test questions were examined in order to assess the quality and quantity of the items and the test as a whole. Test items were assessed and screened on the basis of values of difficulty indices *(Di)* and discrimination power *(Dp)*. Total 203 items were finalised for tools for assessing learner's attainment in science (TALS). Test validity was assessed applying Cronbach alpha separately to all tests. Statistical analysis was done using SPSS version 24 (IBM free software trial version). The tool has been developed for the first time for assessing learners' attainment in science for conceptual understanding among learners of tribal areas of Chhattisgarh. The tool will be utilised for baseline study for measurement of learning gaps and designing the interventions for improvement in learning and teaching of science.

Keywords: Climate Change, Knowledge, Beliefs, Teachers, Predictor, Professional Development, Training

Introduction

Science education is a part of mainstream education till secondary stage. After primary education learners are exposed to science as a separate discipline at the elementary stage in the present curriculum (NCERT, 2017). The present curriculum aims to achieve minimum scientific literacy for every citizen of India in general, that is required for leading life in a meaningful manner free from myth and taboos (NCERT, 2018). Quality teaching and learning specifically in science is crucial for developing scientific literacy in the young generation and also for improving economic productivity for sustainable development (Tripathi *et al,* 2017). Present work is continuation of our earlier work on finalisation of the tools which were designed to find out the gaps in science teaching and learning in scheduled tribes-dominated districts of Chhattisgarh in which three tools including attitude scale for learning science, classroom observation schedule, school observation schedule were finalised (Tripathi *et al,* 2017).

However, the tool for assessing learners' attainment in learning science (TALS) for Classes VI, VII and VIII after applying Kuder Richardson formula 20 (KR 20) in SPSS version 24 (IBM), the individual item analysis was suggested (Tripathi *et al.,* 2017). Thus, to finalise the TALS for error-free measurement of conceptual understanding of science and to find out gaps where students find difficulty to build-up concepts of science, item analysis was applied and individual items in the tool were refined and finetuned.

The tool for assessing learners' attainment in learning science (TALS) for Classes VI, VII and VIII, consisted of a pool of multiple choice questions (MCQs) based on certain criteria including assessment of learning outcomes in science defined by NCERT (NCERT, 2018). Due to many reasons, MCQs are the best items suitable for developing a research tool (Swanson *et al.,* 2006; 2005) for gathering the factual data regarding any subject specific knowledge. Firstly, the tools having MCQs take lesser time and easy for administration even in huge samples in large scale assessment (Haladyna, 2004; Haladyna and Rodriguez, 2013; Rodriguez, 2016). Secondly, they enable measurement of diversified knowledge, skills and competencies across various domains (Haladyna, 2004; Downing, 2006; Popham, 2008) and thirdly, multiple choice questions are the items that are comparatively easier to analyse for drawing conclusions (OECD, 1999), and even more easier with the help of technology intervention and software programming. The crucial thing of preparing tests with multiple choice questions is to construct good items. The development of multiple choice items is highly challenging specifically development of distractors (Miriam Fuhrman, 1996). This requires a good understanding of the objectives of assessment, having good skills in writing the items and an excellent knowledge of the content. (Haladyna, 2004; Walsh, 2008).

Usually two approaches are being used for analysis of an MCQ for taking a decision regarding inclusion or dropping or revision and refinement of items to make tools for effective, accurate and fair assessment (Ary *et al.,* 2002):

1. The Classical Test Theory is based on the item facility index derived by the percentage of students that correctly answered the item and the discrimination index based on the point between students' performance on individual item and total test score.
2. The Item Response Theory (IRT) that describes both item statistics and student's ability with the assumption of correlation between the score on a single item and overall test performance.

Present work is the continuation of our earlier work. The data was collected using the tool for assessing learners' attainment in learning science (TALS) for Classes VI, VII and VIII (Tripathi *et al.,* 2017). After expert opinion and field try out, the tool was determined for the internal consistency by using Kuder Richardson 20 formula. Since the KR-20 values were less than 0.65, it was suggested to analyse individual items by applying IRT for their refinement and improvement (Tripathi *et al.,* 2017). In the present work IRT was applied using difficulty index and discrimination power for a systematic refinement and finalisation of assessment tool (TALS). Data collection was done for the measurement of implementation of recommendations of National Curriculum

Framework (2005) in learning of science in selected rural schools of two scheduled tribal dominated districts of Chhattisgarh State of India. So that, the baseline status of teaching and learning of science in the schools of scheduled tribes dominated districts of the state Chhattisgarh may be assessed and the learning gaps may be identified, intervention may be planned and tried out and a framework may be developed for designing an in-service teacher training programme.

Methodology

Theoretical Background

IRT have now been widely accepted and universally used in the educational research world over in the area of psychometrics and educational measurements (Carlson and von Davier, 2017). Analysis of an individual item using IRT is a statistical technique useful for assessment of items used for objective evaluation of students (Carlson and von Davier, 2017). Item analysis is used for selecting and rejecting the items of any tool on the basis of their difficulty index and discrimination index (Suruchi and Rama, 2015; Lange *et al.,* 1967). It is an important technique to increase the effectiveness of an assessment tool (Gronlund, 1998). It is a process which examines student's responses to individual test items to assess the quality of these items and quality of assessment as a whole. Moreover, in item analysis both students' answers and test questions are examined in order to assess the quality and quantity of the items and the test as a whole. In item analysis an item is assessed after its administration to a population sample (Remmers *et al.,* 1967). The quality of an assessment tool is influenced by quality of individual items of that particular assessment tool (Shrama, 2000 and Freeman, 1962).

IRT has now been used for assessing the tools consisting of MCQs in various subject domains. Use of IRT for validation and analysis of assessment tools in science for refinement of its Multiple Choice Questions have also been reported (OECD, 2016; OECD, 1999). The assessment tool consisting of MCQs for assessing learning outcomes attained by medical students were analysed using IRT by Rao *et al.* (2016). A study was conducted by PhD students of the University of Genoa using item analyses to improve test items that provided valuable information to the middle stage teachers for further modification of items after analysis of students' response (Siri and Freddano, 2011). Item analysis was used for finalisation of tool for a students' attainment on a test, recognition of the items which could not perform when administered in a population and pinpointedly provided suggestion for improvement for better performance of the tool for error free assessment of students' learning outcomes (Gajjar *et al.* (2014); Vanderbilt *et al.* (2013).

Research Procedure

Research methodology was followed as depicted in Figure 1. There were nine Sets developed under this tool on the basis of three thematic groups for Classes VI, VII and VIII as mentioned by Tripathi *et al.,* 2017. Accordingly draft tools for TALS were administered in selected schools in Raipur Chhattisgarh (Tripathi *et al.,* 2017). The present study focuses on the IRT for finalisation of draft tools. Selection

and elimination of items was based on acceptable values of Difficulty index *Di* and Discrimination power *Dp* of individual items.

four schools of Raipur, Chhattisgarh. Details of research procedure and approaches used are presented in Table 1

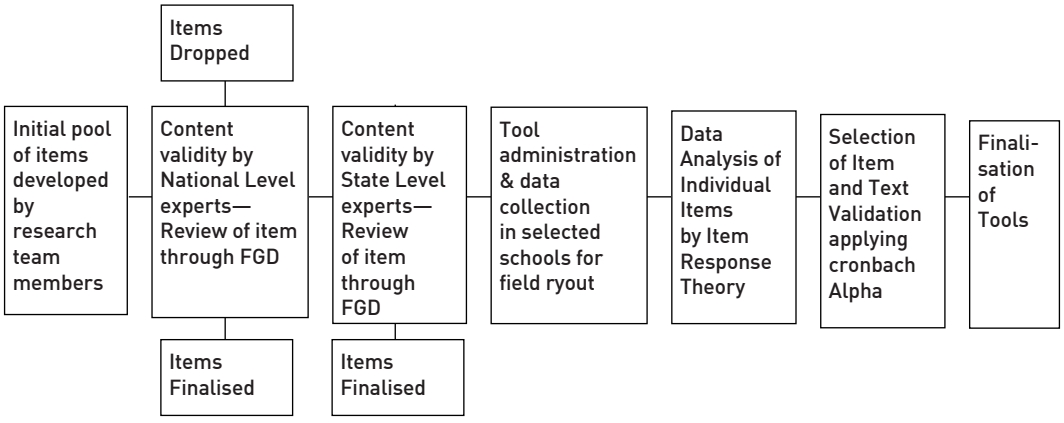

Figure 1: Illustration Showing Systematic Research Methodology Adopted in the Research Study for Finalisation of Tool

This paper is limited to the analysis of individual items using IRT (Item Difficulty Index and Item Discrimination power), their modification and refinement by language interventions based on IRT analysis and finalisation of tool after the validation and internal consistency analysis applying Cronbach's Alpha (Cronbach, 1951).

Finalisation of TALS is necessary for administration in the selected population of learners of Class VI, VII and VIII in select tribal dominated schools in two districts Korba and Kanker of Chhattisgarh, India, to find out the learning gaps in science and accordingly provide the interventions to meet out the objectives of the study.

Study Design

Study design opted in the research broadly was analytical. It was an observational design and a cross sectional study was carried out in

Sample and Sampling Techniques

Sample of the present study was drawn by using multistage sampling techniques in identified schools at Raipur, Chhattisgarh. The technique of sampling was purposive and Government schools at Raipur were chosen where the permission of piloting the instrument was obtained through SCERT, Chhattisgarh. Population of interest was students at middle stage. Total 516 students; 164 from Class VI, 168 from Class VII and 172 from Class VIII participated and filled up the tools for Assessing Learners Attainment in Science (TALS) as mentioned in Figure:2 for assessing attainment of learners in science for Classes VI, VII and VIII (TALS).

Data Analysis

The data obtained after field administration was entered in Microsoft Excel 2010 and analysed using IBM SPSS version 20. Item

Table 1: Study Design and Approach Used for Finalisation of TALS

Research Procedure	Study Design (Approach Used)
Pilot Test and Administration of Assessment Tools for Data Collection	Quantitative Survey Method
Analysis of Data	Using Item Response Theory (IRT); Difficulty Index (Di) and Discrimination Power (Dp)
Selection, Modification, Elimination of Items and Construction of Items	Expert Opinion and Focus Group Discussion on Analysed Data Using IRT
Validation of Items and Internal Consistency of Tools	By Calculating Cronbach's Alpha Using SPSS Version 20
Finalisation of Tool (TALS)	Expert's Opinion and Focus Group Discussion by Analysis of Data and Validation of Tool Items

parameter considered for analysis were Difficulty index (*Di*) and Discrimination power (*Dp*) in the present study. Difficulty index (*Di*) and Discrimination power (*Dp*) for individual items were calculated and analysed for selection and elimination of items.

Difficulty Index (*Di*) is percentage of students who have chosen the correct answer in case of multiple-choice questions or objective type item. Difficulty index of a test item is calculated using formula as mention below:

Di = HP+LP/N x 100

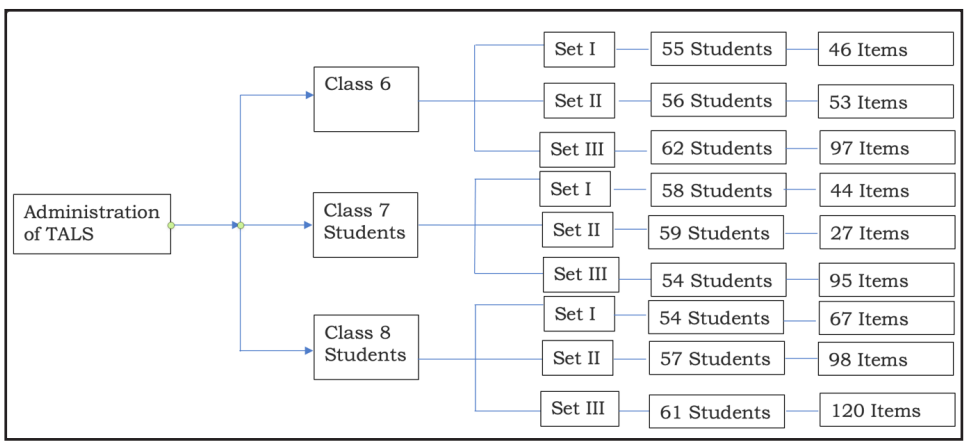

Fig: 2 Illustration Showing Number of Respondents and Number of Items Administered Class-wise and Theme-wise

Here, HP—Total number of students chosen correct answer in high achieving group

LP—Total number of students chosen correct answer in low achiever group

N—Total number of respondents

Analysis of MCQs was carried out on the basis of the values obtained by arranging marks of all students class wise and set wise separately in descending order from highest to lowest score and categorised in three groups 30 per cent high achievers, 40 per cent middle achievers and 30 per cent low achievers. On the basis of difficulty index, items were categorised as given in Table 2 and action was taken accordingly.

Dp may be interpreted as more students in lower group are answering correctly than high achieving group. Dp is calculated by method followed by Kelley *et al.* (2002) using following formula:

$$Dp = 2 \times (HG - LG)/N$$

Here, H—Total number of Students chosen correct answer in high achieving group

L—Total number of Students chosen correct answer in low achieving group

N—Total number of respondents

Table 2: Acceptable Values of Di for Analsis of Items

S. No.	Difficulty Index (Di) Value	Difficulty Status	Action to be Taken
1.	Less than 0.30	Difficult	Modify
2.	Between 0.30 to 0.40	Moderate	Accept
3.	More than 0.60	Easy	Modify

Discrimination Power (*Dp*): Discrimination power is capability of an individual item to discriminate among high and low achievers. It ranges from -1 to +1. If the Dp of an item is at the higher side, the item is more efficient to discriminate between high and low achievers. Item with lower Dp are not competent enough to discriminate between high and low achievers. Negative value of

Field Test and Item Reduction

Selection, modification and removal of an item was based on the acceptable values of difficulty index *(Di)* and discrimination power (Kelley, 1939) of an individual item. Difficulty index and discrimination power *(Dp)* of an individual item was calculated statistically using SPSS version 20 in this study.

Table 3: Status of Discrimination power (Dp) and action to be taken on an MCQ based on Discrimination power (Dp)

S. No.	Discrimination Power (Dp) Value	Discrimination Status	Action to be Taken
1.	Negative value	Item cannot discriminate between high achiever and low achiever	Such items were dropped
2.	Less than 0 to 0.24	Item cannot discriminate between high achievers and low achiever	Such items were accepted with modifications
3.	Between 0.25 to 0.39	Item is efficient to discriminate between high and low achievers	Such items were accepted with little modifications
4.	More than 0.40 to 1	Item is more efficient to discriminate between high and low achievers	Such items were selected as their original form

A rigorous analysis by the expert group followed by a Focus Group Discussion was held for emerging out experts' opinion to emerge out with consensus for refinement of individual items. This way the items were finalised.

Test validity by Cronbach's Alpha

All nine sets of the finalised class-wise and set-wise tools having selected items and the items recommended for modifications due to very low or very high difficulty index or due to low discrimination power between low and high achievers were analysed by applying Cronbach's Alpha measurement for internal consistency (Taber, 2018, Cronbach, 1951).

Results and Discussion

Results of all nine of questionnaires of TALS have been presented after applying Item response theory (IRT). Difficulty indices *(Di)* and Discrimination power *(Dp)* have been considered for analysis.

Analysis of items based on Difficulty indices (*Di*) and Discrimination power (*Dp*)

Results of interpretation of values of difficulty indices (*Di*) and discrimination powers (*Dp*) for each of the selected items for three Sets of questionnaires each for Classes VI, VII and VIII are presented in Table 4, 5 and 6, respectively. The items were selected to be retained or for modifications based on the values of difficulty indices and discrimination powers for finalising the tool for each of the Classes VI, VII and VIII.

The Table 4 reveals results of Set I, Set II and Set III for Class VI. In Set I, out of 46 items administered only 20 items were selected. Based on *Di* and *Dp* only 7 items were retained but 13 were modified keeping in view maintaining the ratio of easy, moderate and difficult items as 25:50:25 respectively. Out of 13 items, 5 were too difficult as the *Di* was less than 0.3, while 7 items were too easy as the *Di* is more than 0.5. Two items were modified keeping in view the values of *Di* and *Dp*. There

Table 4: Item Difficulty and Discrimination Power Class VI

Set I									
S. No.	Item	Item Discriminati-on Power *(Dp)*	Difficulty Index *(Di)*	Item Selected (S)/ Modified (M)	S. No.	Item	Item Discriminati-on Power *(Dp)*	Difficulty Index *(Di)*	Item Selected (S)/ Modified (M)
1	item:1	0.47	0.33	S	11	item:17	0.73	0.35	S
2	Item:2	0.53	0.42	S	12	item:18	0.80	0.55	M
3	item:3	0.80	0.40	S	13	item:32	0.33	0.25	M
4	item:4	0.47	0.22	M	14	item:36	0.60	0.58	M
5	item:5	0.53	0.22	M	15	item:37	0.53	0.54	M
6	item:7	0.60	0.67	M	16	item:38	0.60	0.38	S
7	item:10	0.73	0.69	M	17	item:39	0.40	0.36	S
8	item:14	0.40	0.25	M	18	item:40	0.60	0.51	M
9	item:15	0.73	0.64	M	19	item:41	0.33	0.65	M
10	item:16	0.67	0.27	M	20	item:43	0.67	0.45	S
Set II									
1	item:1	0.60	0.46	S	14	item:23	0.53	0.38	S
2	item:5	0.53	0.61	M	15	item:25	0.47	0.32	M
3	item:8	0.53	0.57	M	16	item:28	0.20	0.23	M
4	item:10	0.80	0.23	M	17	item:29	0.73	0.46	M
5	item:11	0.47	0.48	S	18	item:32	0.40	0.23	M
6	item:14	0.33	0.20	M	19	item:33	0.67	0.50	M
7	item:15	0.73	0.66	M	20	item:41	0.40	0.27	M
8	item:16	0.20	0.38	M	21	item:45	0.47	0.52	M
9	item:17	0.60	0.36	M	22	item:46	0.33	0.29	M
10	item:18	0.53	0.27	M	23	item:50	0.53	0.71	M
11	item:19	0.47	0.27	M	24	item:51	0.40	0.23	M
12	item:20	0.47	0.27	M	25	item:52	0.40	0.20	M
13	item:21	0.47	0.29	M	26	item:53	0.67	0.61	S
Set III									
1	item:1	0.59	0.42	S	18	item:30	0.24	0.21	M

2	Item:2	0.47	0.34	S	19	item:31	0.65	0.35	S
3	item:3	0.41	0.34	S	20	item:40	0.53	0.24	M
4	item:4	0.53	0.40	S	21	item:49	0.53	0.23	M
5	item:5	0.29	0.34	M	22	item:52	0.35	0.24	M
6	item:6	0.35	0.31	M	23	item:75	0.41	0.26	M
7	item:9	0.35	0.24	M	24	item:79	0.71	0.48	S
8	item:11	0.29	0.34	M	25	item:80	0.06	0.48	M
9	item:12	0.24	0.32	M	26	item:81	0.41	0.42	S
10	item:13	0.24	0.44	M	27	item:82	0.41	0.40	S
11	item:14	0.24	0.29	M	28	item:83	0.41	0.48	S
12	item:22	0.41	0.48	S	29	item:84	0.71	0.47	S
13	item:23	0.18	0.32	M	30	item:86	0.29	0.34	M
14	item:24	0.53	0.31	S	31	item:90	0.53	0.40	S
15	item:25	0.41	0.27	M	32	item:94	0.47	0.61	M
16	item:26	0.24	0.27	M	33	item:97	0.59	0.60	M
17	item:29	0.24	0.29	M	34	03 New items			

were 53 items initially for piloting in Set II from thematic area II. Out of which 27 items were dropped due to very low values of *Di* and *Dp*. Number of items retained as it is due to acceptable values of both *Di* and *Dp*. Rest of the 22 items were modified to maintain the ratio of easy, moderate and difficult items to 25:50:25, respectively based on *Di* and *Dp* values. A total of 97 items were administered among the population of Class VI students from thematic area III denoted by Set III. The interpretation of *Di* and *Dp* of 33 selected items are presented in Set III, Table 3. Out of 33 items, 13 items were retained without modifications, while 20 items were modified based on *Di* and *Dp* values. However, it was found that some important concepts were not included in these 33 items and with the help of expert opinion 03 more items were developed afresh. This way 36 items were finalised for inclusion in Set III of the tool for thematic area III for Class VI.

The Table 5 depicts the results of interpretation of *Di* and *Dp* after piloting of tool among students of Class VII. In Set I from thematic area 1, 44 items were screened for piloting after obtaining expert opinion. On the basis of IRT, the two considered parameters *Di* and *Dp*, 16 items were modified and 4 items were retained in the tool for the thematic area 1 for Class VII. In Set II for thematic area II for Class VII, 27 items were screened for administration among the population. In Set III, for thematic area III for Class VII, total 95 items were administered and on the basis of interpretation of *Di* and *Dp*, total 64 items were dropped. Total number of items retained were 8 and 23 items were modified on the basis of analysis of *Di* and *Dp* values.

Table 5: Item Difficulty index (Di) and Discrimination Power (Dp) of Selected Items for Class VII

S. No.	Item	Item Discriminati-on Power *(Dp)*	Difficulty Index *(Di)*	Item selected (S)/ Modified (M)	S. No.	Item	Item Discriminati-on Power *(Dp)*	Difficulty Index *(Di)*	Item selected (S)/ Modified (M)
	Set I								
1	item:1	0.25	0.28	M	11	item:28	0.19	0.45	M
2	Item:2	0.81	0.47	S	12	item:29	0.44	0.31	S
3	item:8	0.75	0.50	M	13	item:33	0.25	0.28	M
4	item:9	0.69	0.52	M	14	item:34	0.69	0.26	M
5	item:11	0.19	0.67	M	15	item:35	0.38	0.28	M
6	item:18	0.31	0.59	M	16	item:37	0.50	0.22	M
7	item:19	0.19	0.31	M	17	item:40	0.38	0.38	M
8	item:20	0.56	0.59	M	18	item:42	0.38	0.40	M
9	item:21	0.38	0.48	S	19	item:43	0.81	0.55	M
10	item:25	0.44	0.26	M	20	item:44	0.69	0.48	S
Set II									
1	item:1	0.31	0.37	M	9	item:12	0.13	0.66	M
2	Item:2	0.50	0.53	M	10	item:15	0.44	0.31	S
3	item:3	0.63	0.44	S	11	item:18	0.38	0.20	M
4	item:5	0.63	0.46	S	12	item:19	0.44	0.37	S
5	item:7	0.25	0.32	M	13	item:21	0.56	0.15	M
6	item:9	0.88	0.39	S	14	item:26	0.44	0.29	M
7	item:10	0.63	0.37	S	15	item:27	0.69	0.51	M
8	item:11	0.44	0.44	S					
Set III									
1	item:1	0.53	0.44	S	17	item:37	0.60	0.31	M
2	item:3	0.13	0.31	M	18	item:40	0.67	0.26	M
3	item:4	0.33	0.30	M	19	item:41	0.47	0.27	M
4	item:5	0.20	0.44	M	20	item:58	0.27	0.44	M
5	item:7	0.73	0.59	M	21	item:59	0.47	0.44	S
6	item:8	0.33	0.72	M	22	item:60	0.33	0.46	S

7	item:9	0.60	0.61	M	23	item:61	0.53	0.43	S
8	item:10	0.53	0.67	M	24	item:62	1.0	0.44	S
9	item:24	0.20	0.30	M	25	item:64	0.47	0.30	S
10	item:26	0.20	0.24	M	26	item:67	0.27	0.31	M
11	item:29	0.53	0.35	S	27	item:69	0.27	0.22	M
12	item:31	0.13	0.33	M	28	item:70	0.20	0.26	M
13	item:32	0.33	0.43	M	29	item:89	0.27	0.52	M
14	item:34	0.60	0.30	S	30	item:90	0.33	0.54	M
15	item:35	0.60	0.20	M	31	item:95	0.13	0.44	M
16	item:36	0.60	0.30	M					

Table 6: Item Difficulty index (Di) and Discrimination Power (Dp) for Selected Items for Class VIII

S. No.	Item	Item Discriminat-on Index	Difficulty Index	Item selected/ dropped/ modified	S. No.	Item	Item Discriminati-on Index	Difficulty Index	Item selected(S)/ modified (M)
Set I									
1	item:4	0.40	0.76	M	11	item:34	0.53	0.67	M
2	item:5	0.20	0.65	M	12	item:37	0.40	0.56	M
3	Item:14	0.73	0.46	S	13	item:38	0.47	0.31	S
4	Item :15	0.80	0.57	M	14	Item:39	0.33	0.59	M
5	item:16	0.80	0.56	M	15	item:40	0.53	0.41	S
6	item:17	0.73	0.33	S	16	item:41	0.53	0.52	M
7	item:18	0.47	0.31	S	17	item:42	0.40	0.56	M
8	Item:23	0.40	0.33	S	18	item:45	0.47	0.37	S
9	item:25	0.93	0.59	M	19	item:54	0.67	0.43	S
10	item:29	0.40	0.48	S	20	item:64	0.40	0.31	M
Set II									
1	item:10	0.53	0.56	M	11	item:67	0.40	0.56	M
2	item:12	0.53	0.35	S	12	item:69	0.47	0.54	M
3	item:13	0.40	0.30	S	13	item:71	0.47	0.49	S
4	item:24	0.60	0.37	S	14	item:77	0.53	0.42	S
5	item:25	0.67	0.42	S	15	item:84	0.80	0.63	M

6	item:36	0.47	0.46	S	16	item:93	0.67	0.54	M
7	item:37	0.53	0.37	S	17	item:94	0.47	0.53	M
8	Item: 39	0.47	0.44	S	18	item:95	0.60	0.54	M
9	item:52	0.33*	0.63	M	19	item:96	0.47	0.39	S
10	item:64	0.47	0.67	S	20	item:97	0.60	0.61*	M
Set III									
1	item:1	0.65	0.34	S	9	item:41	0.41	0.31	S
2	item:12	0.35	0.46	M	10	item:70	0.35	0.31	M
3	item:13	0.53	0.46	S	11	item:81	0.65	0.33	S
4	item:14	0.41	0.48	S	12	item:85	0.47	0.39	S
5	item:15	0.59	0.41	S	13	item: 115	0.29	0.51	M
6	item:17	0.53	0.31	S	14	item: 116	0.41	0.57	M
7	item:20	0.59	0.49	S	15	item: 118	0.29	0.56	M
8	item:23	0.47	0.33	S					

Comparison of screened test items on the basis of their *Di* and *Dp* interpretation is presented in Table 6 for Class VIII for three thematic areas as Set I, Set II and Set III. Set I for thematic area 1, 64 items were administered to a population of students of Class VIII. The data obtained were analysed on the basis of *Di* and *Dp* and total 47 items were dropped. Out of total 20 items, 9 items were retained without modifications and 11 items were modified. Likewise, For Set II, 98 items were administered and interpretation of *Di* and *Dp* suggested to select only 20 items, out of which, 11 item to retain without modifications and 9 items to be modified as per the requirement to make them easy, moderate or difficult. A total of 120 items for Set III were administered among the students of Class VIII and on the basis of Di and Dp interpretations 10 items were selected and 5 items were modified to be part of the tool for Set III of Class VIII.

Finalisation of Tools

Individual items were screened by applying IRT and analysis of the data based on *Di* and *Dp*. In this tool TALS three Sets each for Classes VI, VII and VIII finalised. Total number of items finalised for inclusion in all nine Sets class-wise and theme-wise have shown in Table 7 for Class VI, Set I, Set II and Set III for three thematic areas, total 20, 26 and 36 items were finalised, respectively. Likewise, for Class VII, Set I, Set II and Set III for three thematic areas, total 20, 15 and 31 items were finalised, respectively. For Class VIII, Set I, Set II and Set III for three thematic areas, total 20, 20 and items 15 were finalised respectively. Total number of 82, 66 and 55 items were finalised in Classes VI, VII and VIII, respectively. Total 203 items were developed and finalised theme-wise for assessing middle stage learners. Five items from each of the sets of each of the classes were identified and added to other sets. For example, for Class VI Set I,

Table 7: Summary of Total Number of Items of Finalised Class-wise and Theme-wise

Dimension		Total Number of Items	Number of Items Dropped	Total Number of Items Selected		Items Finalised	Items in Questionnaire (05 Items Each from Other Sets)
				Retained	Modified		
Class 6	Set I	46	26	07	13	20	30
	Set II	53	27	04	22	26	36
	Set III	97	61	13	20 and 03 developed new	36	46
Class 7	Set I	44	24	04	16	20	30
	Set II	27	12	07	08	15	25
	Set III	95	64	08	23	31	41
Class 8	Set I	67	47	09	11	20	30
	Set II	98	78	11	09	20	30
	Set III	120	105	10	05	15	25
Total		647	444	73	127 and 03 developed new	203	293

Set II and Set III, five items from each Set were identified and added to other Sets, that means, in Set I five items from Set II and five item from Set III; in Set II, five items from Set I and five items from Set III and in Set III, five items from Set I and five items from Set II were added and finalised the tools. Therefore, in finalised tool in each Set 15 items were common for each class.

Results of Data Analysis and Finalisation of Items

Cronbach's Alpha for Internal Consistency of Test:

Cronbach's alpha coefficient for each dimension was calculated using SPSS version 24 (IBM) for internal consistency reliability analysis. Acceptable range was ≥ 0.60 in this study according to several other studies (Ware *et al.*, 1980 and Alotaibi; Youssef, 2013).

The data was analysed on the basis of the statistical properties that is difficulty index (*Di*) and discrimination power (*Dp*) of items for further modifications or for its elimination if not appropriate by researchers. Later, along with a group of experts in science and science education; including experts from national level and state level experts from Chhattisgarh state of India; the items in each tool of the instrument were finalised after review as per the analysis.

The present study testifies and describes the process and outcomes of developing an instrument to assess learners' attainment in learning of science at middle (IRT). The

instrument uses response of population of students at middle stage, Classes VI, VII and VIII for analysing the individual items based on the status of difficulty index (*Di*) and discrimination power (*Dp*) by Kelley's method (1939). The study has important implications for development of appropriate tools for assessing knowledge and competencies in science in the learners of scheduled tribe dominated areas. Such assessment tools may help identifying learning gaps and provide insights for interventions in the teaching and learning process. Earlier no research efforts were made for development of assessment tools for assessment of learners' attainment in subject domain of science, especially in tribal dominated areas. Therefore, it is crucial to develop an error-free and applicable tool that can provide an insight and a clearer understanding of needs in teaching learning process and guides content that is relevant for curriculum implications in future. Such tools may provide insights for multifaceted teacher education programs in science teaching. The outcomes of piloting the instrument and the analysis of individual items indicates that some items are easy and some are difficult while in case of others, no correct answer was obtained from any of the respondents. After analysing the internal consistency using Cronbach alfa (Cronbach, 1951), the finalised instrument can be used in its current state for assessing learners of tribal dominated areas of Kanker and Korba districts of Chhattisgarh for obtaining learning gaps and providing interventions for improvement in teaching and learning process of science. However, more statistical methods may be applied for bringing more accurate items for meeting out the objectives in the instruments.

Limitations

The study provides an instrument for assessment of learners' attainment in learning science only for learners of some tribal dominated districts of Chhattisgarh near Raipur. For other areas including tribal dominated areas, the instrument needs to be modified keeping in view other interacting factors, such as, language efficiencies, cultural inputs, contextualisation, environment and surroundings, existing curricular material and so on. Distractor analysis have not been carried out and it can be done for further modifications and refinement in the options of multiple choice items.

Authors' Contributions

First author wrote the manuscript of the paper and carried out statistical analysis using SPSS free version 24 of IBM with the help of second author. Third, fourth and fifth authors analysed, reviewed and modified the items of the tools along with the first author and finalised the tools.

stage in schools of tribal dominated areas of Chhattisgarh using a mixed methods approach. Since literature regarding the development of such kinds of tools and IRT analysis were not available for assessing science learning, however, literature were available regarding development of such tools with MCQs in different subject domains and finalisation individual items by IRT analysis (Carlson and von Davier, 2017; Lange; Gronlund, 1993; Remmers *et al.,* 1967; Sharma, 2000; Freeman, 1962; OECD 2016; Siri and Freddano, 2011; Rao *et al.*. (2016); Gajjar, *et al.* (2014); Vanderbilt, *et al.*. (2013).

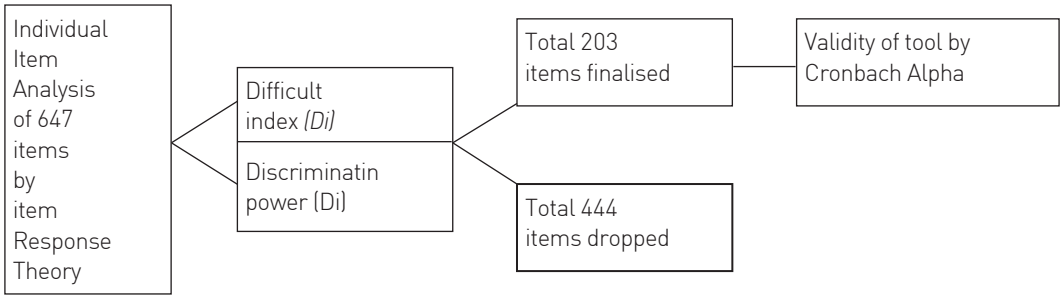

Fig: 3 Finalisation of Items for TALS

Table 8: Internal Consistency of class-wise and theme-wise tools using Cronbach's Alpha

Dimensions		Number of Items Selected Based on *Di* and *Dp*	Cronbach's Alpha
Class VI	Set I (Thematic group I)	20	0.875
	Set II (Thematic group II)	26	0.816
	Set III (Thematic group III)	33	0.798
Class VII	Set I (Thematic group I)	20	0.720
	Set II (Thematic group II)	15	0.607
	Set III (Thematic group III)	31	0.753
Class VIII	Set I (Thematic group I)	20	0.827
	Set II (Thematic group II)	20	0.837
	Set III (Thematic group III)	15	0.700

Conclusion

The present study includes the process of developing an instrument by analysing each of the items applying Item Response Theory finalised the tools.

References

ALOTAIBI G, A. YOUSSEF DEVELOPMENT OF AN ASSESSMENT TOOL TO MEASURE STUDENTS' PERCEPTIONS OF RESPIRATORY CARE EDUCATION PROGRAMS: ITEM GENERATION, ITEM REDUCTION AND PRELIMINARY VALIDATION. *J FAMILY COMMUNITY MED.* 2013 May. Vol. 20, No.2. pp. 116–22. doi: 10.4103/2230-8229.114770. PMID: 23983564; PMCID: PMC3748646.

ARY D., L.C. JACOBS, A. RAZAVIEH. *INTRODUCTION TO RESEARCH IN EDUCATION,* 6th edition, Wadsworth, California, 2002.

CARLSON, J.E., M. VON DAVIER. 2017. Item Response Theory. In: R. Bennett, M. von Davier, (Eds) *Advancing Human Assessment. Methodology of Educational Measurement and Assessment.* Springer, Cham. https://doi.org/10.1007/978-3-319-58689-2_5

CRONBACH, L. J. 1951. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika.* Vol. 16, No.3. pp. 297–334. doi:10.1007/bf02310555

DOWNING, S. M. 2006. 'SELECTED-RESPONSE ITEM FORMATS IN TEST DEVELOPMENT,' IN T. M.HALADYNA AND S. M. DOWNING. HANDBOOK OF TEST DEVELOPMENT, (EDS). DIDCOT: TAYLOR & FRANCIS GROUP. pp. 287–301.

FREEMAN FRANK S. 1962. *Theory and Practice of Psychological Testing.* Oxford and Ibh publishing New Delhi.

GAJJAR, S., R. SHARMA, P. KUMAR, *et al.* 2014. Item and Test Analysis to Identify Quality Multiple Choice Questions (MCQs) from an Assessment of Medical Students of Ahmedabad, Gujarat. Indian Journal of Community Medicine, Vol. 39, No. 1. pp. 17–20.

GRONLUND, N. E. 1998. *Assessment of Student Achievement*. 6th Edition. Allyn and Bacon, Boston.

HALADYNA, T. M. 2004. Developing and Validating Multiple-Choice Test Items, 3rd Edition: Erlbaum, Mahwah, NJ.

HALADYNA, T. M. AND M. C. RODRIGUEZ. 2013. Developing and Validating Test Items. New York, NY: Routledge. doi: 10.4324/9780203850381

KELLEY T., R. EBEL AND J. M. LINACRE. 2002. Item *Discrimination Indices*. *Rasch Measurement Transactions*. Vol. 16, No. 3. pp. 883–884.

KELLEY T. L. 1939. The Selection of Upper and Lower Groups for the Validation of Test Items. *Journal of Educational Psychology.* Vol. 30. pp. 17–24. DOI: http://dx.doi.org/10.1037/h0057123. Reprinted in A. Ward, H. W. Stoker, and M. Murray-Ward. 1996. Educational Measurement: Origin, Theories and Explications. Vol II: Theories and applications. University Press of America *(11) (PDF) Generalized Discrimination Index.* https://www.researchgate.net/publication/327631668_Generalized_Discrimination_Index .

Kumar D., R. Jaipurkar, A. Shekhar, G. Sikri and V. Srinivas Item analysis of multiple choice questions: A quality assurance test for an assessment tool. Med Journal Armed Forces India. 2021 Feb;77(Suppl 1):S85-S89. doi: 10.1016/j.mjafi.2020.11.007. Epub 2021 Feb 2. PMID: 33612937; PMCID: PMC7873707.

Lange, A., I.J. Lehmann and W.A. Mehrens. 1967. Using Item Analysis to Improve Tests. *Journal of Educational Measurement.* Vol. 4, No. 2. pp. 65–68. http://www.jstor.org/stable/1434299.

Miriam Fuhrman. 1996. Developing Good Multiple-Choice Tests and Test Questions. *Journal of Geoscience Education.* Vol. 44, No. 4. pp. 379–384. DOI: 10.5408/1089-9995-44.4.379 https://www.tandfonline.com/action/showCitFormats?doi=10.5408%2F1089-9995-44.4.379

NCERT. 2018. Report (PAC 6.02, (2015-2017)), Promoting Quality Learning in Science at Upper Primary Level in Scheduled Tribes (ST) dominated Districts of Chhattisgarh: A Framework.

NCERT. 2017. Learning Outcomes at the Elementary Stage. pp. 110.

NCERT. 2005. National Curriculum Framework. pp. 158. NCERT, New Delhi.

OECD. 2016. *Education at a Glance 2016: OECD Indicators*. OECD Publishing. Paris, https://doi.org/10.1787/eag-2016-en.

OECD. 1999. Measuring Student Knowledge and Skills: A New Framework for Assessment. Organisation for Economic Cooperation and Development, Paris. https://www.oecd.org/Education/School/Programme for International Student Assessment pisa/33693997.pdf

Popham, W. J. 2008. *Transformative Assessment.* ASCD Alexandria, VA.

Rao C., H. L. Kishan Prasad, K. Sajitha, H. Permi and J. Shetty. 2016. Item Analysis of Multiple Choice Questions: Assessing Anassessment Tool in Medical Students. Int J Educ Psychol Res. Vol .2. pp. 201–204.

Remmers H.H., N.L. Gage and J.I. Rummel. 1967. *A Practical Introduction to Measurement and Evaluation* (2nd Edition). Universal Book Stall Delhi.

Rodriguez, M. C. 2016. 'Selected-Response Item Development,' in S. Lane, M. Raymond and T. Haladyna Handbook of Test Development, 2nd Edition(eds) (Routledge, New York, NY:). pp. 259–273.

Sharma, S.R. 2000. *Modern Teaching Strategies.* Omsons Publications, New Delhi.

Siri A. and M. Freddano. 2011 The Use of Item Analysis for the Improvement of Objective Examinations. *Procedia — Social and Behavioral Sciences*. Vol. 29, No. 3. pp. 188–197 DOI: 10.1016/j.sbspro.2011.11.224

Suruchi and S.R. Rana. *Test item analysis and relationship between difficulty level and discrimination index of test items in an achievement test in biology. Indian J. Res.* 2015. Vol. 3, No. 6. pp. 56–58.

Swanson, D.B., K.Z. Holtzman, K. Albee and B.E. Clauser. 2006. Psychometric Characteristics and Response Times for Content-Parallel Extended-Matching and One-Best-Answer Items in Relation to Number of Options. Acad Med.Vol. 81. pp. 52–55.

Swanson, D.B., K. Z. Holtzman, B.E. Clauser and A.J. Sawhill. 2005. Psychometric Characteristics and Response Times for One-Best-Answer Questions in Relation to Number and Sources of Options. Acad Med. Vol. 80. pp. 93–96.

Taber, K.S. *The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ* Vol. 48. pp. 1273–1296 (2018). https://doi.org/10.1007/s11165-016-9602-2

Tripathi, B K., I. Bhaduri, R. Garg, A. K. Srivastava, D. Kumar, R.W K. Parashar, R. Verma and S Farkya. 2017. Construction and Pilot Test of an Instrument to Measure the Gaps in Teaching-Learning of Science at Middle Stage in Select Tribal Dominated Rural Schools of Chhattisgarh, India. *School Science,* Vol. 55, No.4, pp. 5–21.

Vanderbilt, A. A., M. Feldman and I. K. Wood. 2013. Assessment in Undergraduate Medical Education: A Review of Course Exams. *Medical Education Online.* Vol. 18, No.1. pp. 1–5.

Walsh, K. 2008. *Advice on Writing Multiple Choice Questions (MCQs).* BMJ Careers 2005. Available at http://careers.bmj.com/careers/advice/view-article.html?id=616 , accessed on 23 May 2008.

Ware J.E., R. H. Brook, A.R. Davies, K.N. Williams, A. L. Stewart and W. H. Rogers. *Model of Health and Methodology.* Vol. 1. Santa Monica, CA: RAND Corporation; 1980. Conceptualization and Measurement of Health for Adults in the Health Insurance Study. Doc no R-1987/1-HEW.