



## CONSTRUCTION AND STANDARDIZATION OF THE ACHIEVEMENT TEST IN SCIENCE FOR CLASS VIII STUDENTS

<sup>1</sup>Aju Kurian, & <sup>2</sup>Dr. Zacharias Tirkey

<sup>1</sup>Research Scholar, <sup>2</sup>Assistant Professor,

Department of Education

Assam Don Bosco University, Guwahati, India

**Abstract:** Science is critical to any society's advancement, and if it is deprived of its due role, all civilization comes to a halt. Science prepares us with a diverse set of problem-solving, logical reasoning, and creative and critical thinking abilities. In order to build analytical skills in science, one must use a variety of methodologies. Students can benefit from achievement tests in science since they assess the amount of information or competencies gained by an individual as a result of systematic science instruction. In the present study, a few chosen topics in science that are part class VIII syllabus was used to prepare an achievement test. The researcher, keeping in mind the objectives and content, prepared a preliminary draft having 78 items. After the expert, individual, and group try-outs, only 38 items were included in the final draft. The test-retest method was employed to establish the reliability of the test and the coefficient of correlation was found to be 0.87 which indicated a strong relationship. Expert judgement and content validity judgement were used to establish the validity of the test. The content validity ratio was found to be 0.99.

**Index Terms - Achievement Test, Construction, Standardization, Item Analysis, Reliability, Validity**

### 1. Introduction

Achievement constitutes an individual's learning attainments, his or her accomplishments, success, and proficiency, and it denotes what an individual has achieved after the specific instruction has been imparted (Anastasi, 1968). "A test of educational achievement is one designed to measure knowledge, understanding, or skills in a specified subject or group of subjects (Freeman, 1962, p. 490)". These tests do not only measure the volume of information assimilated or the skills and techniques developed, but also measure educational outcomes such as attitudes and appreciations and the ability to solve problems, draw inferences from subject matters, and apply generalisations to particular situations (Freeman, 1962). A well-designed, valid, and reliable assessment tool aids in the planning of curriculum and the development of appropriate instructional strategies, monitors instructional progress, allows instructors to understand what students learn in a particular course, and provides students with necessary feedback (Singh, 2008; Knight, 2010; Sharma & Poonam, 2017; Aggarwal, 2018; Sharma & Sarita, 2018). It assesses the effectiveness of teachers and evaluates the efficacy of various teaching approaches in a controlled environment.

### 2. The Objectives of the Study

- To construct an achievement test in science (ATS) using selected units of science for grade VIII
- To standardize the achievement test in science by establishing discrimination index, difficulty value, reliability, and validity.

### 3. Target Population

This achievement test is designed for students who are studying in Class VIII

### 4. Method for Creating and Standardizing Science Achievement Tests

The achievement test in science is developed on the basis of the objectives of teaching; knowledge, understanding, application, analysis, and skills in science (Sharma & Poonam, 2017). Before constructing the test, the researcher methodically reviewed a few available achievement tests in science. Following the evaluation, the need was felt to construct an achievement test based on the few selected modules from the most recent ICSE syllabi for science at the eighth-grade level. The following steps are involved in the construction of ATS

- Giving operational definitions
- Planning of the test
- Preparation of the test
- Administration of the test
- Item analysis

f) Standardization of the test

#### 4.1. Giving operational definition

An operational definition is a clear, succinct, and detailed definition of a measurement (Sansanwal, 2020). The terms achievement and science were used in the construction of the present test.

- a) **Achievement:** Achievement means the extent to which a student acquires certain information and demonstrates proficiency in certain skills as a result of systematic instruction imparted on a subject. It was represented by the scores obtained by the students in science.
- b) **Science:** It is a rigorous study and investigation of the structure and behaviour of the physical and natural world that is both intellectual and practical. Science in this study refers to the selected class VIII modules of physics, chemistry, and biology.

#### 4.2. Planning of the Test

Careful and systematic planning is required for the construction and standardization an achievement test. First and foremost, the researcher specified the broad and specific objectives of the test. The objective and content of the test, nature of the test, scoring schemes, number and type of items, length of test, weightage to objectives, content and questions, duration and difficulty level of the test, and marking scheme were taken into consideration while planning for the test (Bhagat & Baliya, 2016; Bala & Singh, 2019).

a) **Objectives of the Test:** Test objectives were defined in behavioural terms, focusing on knowledge, understanding, application, analysis, and skills from selected chapters of the ICSE's science textbook for the VIII class.

**Table 1: Distribution of weightage to objectives (Preliminary Draft)**

Objectives	No. of Questions	Marks	Weightage
Knowledge	26	22	18.33%
Understanding	12	21	17.5 %
Application	22	27	22.5 %
Analysis	13	32	26.67 %
Skill	5	18	15%
<b>Total</b>	<b>78</b>	<b>120</b>	<b>100%</b>

b) **Content of the Test:** The content of the ICSE board's class VIII curriculum was thoroughly examined in order to create ATS. Six chapters, two from each of the three divisions of science (physics, chemistry, and biology), were used to create a well-balanced test.

**Table 2: Subject weightage (Preliminary Draft)**

Content	No. of items	Marks	Weightage
Physics	26	40	33.33%
Chemistry	26	40	33.33%
Biology	26	40	33.33%
<b>Total</b>	<b>78</b>	<b>120</b>	<b>100 %</b>

c) **Size and Form of Test:** The number of items in the test constitutes the size of the test. It is always better to have a quite good number of test items in the initial draft keeping in view the future screening during the process of the successive try-outs and for testing the reliability and the validity of the test. Therefore, the initial draft of the test contained 78 items.

**Table 3: Weightage to the form of questions (Preliminary Draft)**

Form of Questions	No. of Items	Marks	Weightage
Multiple choice	18	9	7.5 %
Fill in the blanks	18	18	15 %
Matching	12	6	5 %
Very short answer	12	24	20 %
Short answer	9	27	22.5 %
Long answer	9	36	30 %
	<b>78</b>	<b>120</b>	<b>100%</b>

d) **Preparation of Blueprint:** Preparation of blueprint is an important phase of planning a test. A blueprint is a detailed plan or framework for any action, and it gives the researcher a bird's-eye view of the complete test and gives the users a foundational understanding of the reasoning behind the test blueprint creation procedure. It serves as the guidelines to guarantee that each objective is given due consideration and also makes sure that the selected units are represented proportionately in the achievement test.

**Table 4: Blueprint of the Preliminary draft of ATS**

Content		Knowledge	Understanding	Application	Analysis	Skill	Total
Physics	Electricity	C (2), F (2)	...	C (1), F (1), M (2), VS (1), L (1)	C (1), VS (1) L (1)	L (1)	14
	Heat Transfer		C (1), F (2), VS (1)	C (1), F (1), M (2)	VS (1), S (2)	S (1)	12

Chemistry	Water	C (3), F (2), VS (1)	VS (1)	M (3), S (1)	S (1)	S (1)	13
	Carbon & its compounds	C (2), F (3), VS (1)	VS (1) L (1)	M (1)	C (1), F (1) L (1)	L (1)	13
Biology	Health & Hygiene	C (2), F (2)	C (1), F (1), VS (1), S (1)	M (2) S (1)	VS (1) S (1)	...	13
	Food Production	C (3), F (3)	VS (1)	M (2), VS (1), L (1)	L (1)	L (1)	13
		<b>26</b>	<b>12</b>	<b>22</b>	<b>13</b>	<b>5</b>	<b>78</b>

- **C:** Choose the correct alternative; **F:** Fill in the blanks; **VS:** Very short answer; **S:** Short answer; **L:** Long Answer

#### 4.3. Preparation of the Test

The preparation of a preliminary draft of a standardized test includes two stages, namely: a) item-writing and b) item-editing.

**Item-Writing:** The investigator prepared a first draft of ATS, which included 78 test items, after thoroughly analyzing the content matter. Efforts were made to ensure that the key objectives were met.

**Item-editing:** Language and subject experts rigorously analyzed and edited items. This aided in correcting ambiguities, verifying the objectives, and putting forward suggestions to add any other relevant test items. A scoring key was also made for the preliminary draft. Proposed items were edited, revised, and rewritten to improve them as per the directions given by the experts. In this way, the first version of the ATS was prepared.

**Method of Scoring:** The scoring was done manually as per the table given. The total score on the test is considered the raw score of the individual student. The raw scores of the students were used in selecting the items for the final administration through an item analysis procedure.

**Table 5: Scoring pattern of ATS**

Form of Question	No. of Items	Marks	Total Marks	Weightage
Multiple Choice	18	½	9	7.5 %
Fill in the blanks	18	1	18	15 %
Matching	12	½	6	5 %
Very short answer	12	2	24	20 %
Short answer	9	3	27	22.5 %
Long answer	9	4	36	30 %
	<b>78</b>		<b>120</b>	<b>100%</b>

#### 4.4. Administration of the Test

**Expert Try-out:** An expert is a person who has knowledge of research methodology, particularly tool development, and knowledge of the subject chosen. The researcher sent the first draught of ATS to 14 experts along with the operational definition, objectives, name of the components, and dimension-wise serial number of items, with the request that each item be carefully scrutinized in the light of the operational definition, dimension, and language. The majority of the items were approved by the experts. A few items were deleted, and a few were modified based on the valuable feedback given by the experts.

**Individual Try-out:** Individual try-outs of ATS were then conducted with the goal of determining the linguistic suitability of the items. This stage allowed the researcher to see if the subjects had grasped the meaning of each item. For the purpose of individual try-outs, 14 students were chosen purposively from the population under study. After obtaining necessary permission from the school management and creating rapport with the students, they were taken one at a time. The researcher described the test's objective and asked each student to read the instructions. All students understood the test instructions. Each item was read by the students, and feedback was recorded immediately. The procedure was repeated until all items were completed. The technique was repeated for the remaining pupils by the researcher. Based on the feedback received from the students, some items were modified and a few were dropped. The test was then used for group try-out.

**Group Try-out:** The primary goal of the group try-out was to conduct item analysis. The difficulty level and discrimination index were used to analyze the items. After completing the expert and individual try-outs, 65 items were selected, and they were arranged in a suitable format with clear instructions. For the purpose of group try-out, a sample size of 100 students was chosen.

#### 4.5. Item Analysis:

After scoring of the test items, item analysis was done by the researcher to assess the effectiveness of all items. Item analysis is a statistical technique employed to assess the suitability of the items on the basis of their difficulty level and discrimination index. The researcher used the item analysis procedure suggested by Ebel (1991). The following steps were followed:

- The scores obtained were arranged in decreasing order.
- The upper and lower groups were identified. The upper group has the highest score of 27% and the lower group has the lowest score of 27%.

- The marks obtained by the first 27 students were considered the students of the upper group, and the marks obtained by the last 27 students were considered the students of the lower group.

**Difficulty Index:** The Difficulty Index of an item is represented by the percentage of examinees who responded correctly to each item. It is calculated by using the formula suggested by Ebel (1991).  $D = (U + L) / 2N$

Where, D = Difficulty Index

U = the number of correct responses in the Upper Group

L = the number of correct responses in the Lower Group

N = the number of individuals in each group

**Table 6: Guidelines for difficulty Index**

Difficulty Index	Item evaluation
0.20 - 0.30	Most difficult
0.30 - 0.40	Difficult
0.40 - 0.60	Moderate difficult
0.60 - 0.70	Easy
0.70 - 0.80	Most easy

**Index of Discrimination:** Index of Discrimination of an item refers to the quality of an item at which it discriminates between examinees with high and low knowledge. It is based on the difference between correct responses in the lower group and upper group. It is measured by using the formula,  $d = (U - L) / N$

Where, d= Discrimination Index

U = the number of correct responses in the Upper Group

L = the number of correct responses in the Lower Group

N = the number of individuals in each group

**Table 7: General guidelines for index of discrimination Ebel (1991, p. 232)**

Index of discrimination	Item evaluation
0.40 and above	Very good items
0.30-0.39	Reasonably good but subject to improvement
0.20-0.29	Marginal items, need improvement
Below 0.19	Poor items, to be rejected or improved by revision

The researcher selected 38 items for the final test. The items whose index of discrimination  $d \geq 0.30$  and the items whose difficulty index 'D' value ranging from 0.30 to 0.70 were included in the final test.

**Table 8: Item-wise number of correct answers, Difficulty Index and Discrimination Index**

Item no.	Top 27%	Bottom 27%	Difficulty Index	Discrimination Index	Remark
<b>A) 1</b>	25	12	0.69	0.48	Selected
2	22	10	0.59	0.44	Selected
3	10	14	0.44	-0.15	Rejected
4	20	8	0.52	0.44	Selected
5	20	4	0.44	0.59	Selected
6	23	12	0.65	0.41	Selected
7	19	10	0.54	0.33	Selected
8	20	9	0.54	0.41	Selected
9	20	8	0.52	0.44	Selected
10	24	12	0.67	0.44	Selected
11	25	10	0.65	0.56	Selected
12	16	7	0.43	0.33	Selected
13	22	11	0.61	0.41	Selected
<b>B) 1</b>	22	7	0.54	0.56	Selected
2	26	22	0.89	0.15	Rejected
3	20	8	0.52	0.44	Selected
4	18	20	0.70	-0.07	Rejected
5	16	4	0.37	0.44	Selected
6	22	9	0.57	0.48	Selected
7	25	24	0.91	0.04	Rejected
8	27	20	0.87	0.26	Rejected
9	19	6	0.46	0.48	Selected
10	8	1	0.17	0.26	Rejected
11	12	6	0.33	0.22	Rejected
12	22	8	0.56	0.52	Selected
13	6	0	0.11	0.22	Rejected
14	25	11	0.67	0.52	Selected
15	14	8	0.41	0.22	Rejected

16	25	23	0.89	0.07	Rejected
17	21	11	0.59	0.37	Selected
<b>C) 1</b>	27	25	0.96	0.07	Rejected
2	25	10	0.65	0.56	Selected
3	20	18	0.70	0.07	Rejected
4	24	23	0.87	0.04	Rejected
5	27	10	0.69	0.63	Selected
6	26	12	0.70	0.52	Selected
7	21	18	0.72	0.11	Rejected
8	25	12	0.69	0.48	Selected
9	26	12	0.70	0.52	Selected
10	27	25	0.96	0.07	Rejected
11	26	22	0.89	0.15	Rejected
12	24	12	0.67	0.44	Selected
<b>D) 1</b>	10	3	0.24	0.26	Rejected
2	20	8	0.52	0.44	Selected
3	24	20	0.81	0.15	Rejected
4	20	5	0.46	0.56	Selected
5	23	10	0.61	0.48	Selected
6	24	19	0.80	0.19	Rejected
7	15	3	0.33	0.44	Selected
8	26	24	0.93	0.07	Rejected
9	18	6	0.44	0.44	Selected
10	25	22	0.87	0.11	Rejected
11	15	2	0.31	0.48	Selected
<b>E) 1</b>	7	1	0.13	0.26	Rejected
2	20	10	0.56	0.37	Selected
3	16	8	0.46	0.26	Rejected
4	21	19	0.74	0.07	Rejected
5	23	20	0.80	0.11	Rejected
6	24	12	0.67	0.44	Selected
7	16	2	0.33	0.52	Selected
<b>F) 1</b>	18	6	0.44	0.44	Selected
2	20	3	0.43	0.63	Selected
3	8	6	0.26	0.07	Rejected
4	16	5	0.39	0.41	Selected
5	4	0	0.07	0.15	Rejected

#### 4.6. Standardization of Achievement Test in Science

**Reliability of the Test:** Reliability refers to the consistency of the scores obtained by the same individual when re-examined with the same test on different occasions or with different sets of equivalent items or under valuable examining conditions (Anastasi, 1951). In order to estimate the reliability of ATS, the investigator used the test-retest method of reliability. For obtaining test-retest reliability, the final version of the test was administered to a sample of 61 students, and the second administration of the test was carried out after a period of 12 days on the same group of students. Scores obtained in both the measures were correlated by using the Pearson product moment method of correlation. The value of the coefficient of reliability was found to be 0.87 which indicated a strong relationship. The result showed a high degree of consistency between the two sets of scores.

**Validity of the Test:** The extent to which a concept is accurately measured in a quantitative study is defined as validity (Heale & Twycross, 2015). Content validity was used in this investigation. Content validity requires the use of recognised subject matter experts to evaluate whether the items assess defined content. Subject matter expert raters were given the items, together with the operational definition, to assess how effectively they assessed the variable for which they were written. To calculate content validity the formula given by C.H. Lawshe (1975) was used, which is as follows:

$$\text{CVR} = (\text{ne}-\text{N}/2) / (\text{N}/2)$$

CVR = Content Validity Ratio

ne = number of subject matter expert raters indicating "essential".

N = total number of subject matter expert raters

This equation yields a value which ranges from -1 to +1; positive values show that at least half the subject matter experts rated the item as "essential." The mean CVR across items may be used as an indicator of overall tool content validity (Sansanal, 2020). The experts agreed with the researcher on the distribution and objective of the content as well as with the scoring scheme. The content validity ratio was found to be 0.99.

#### 5. Conclusion

The objective of this study was to develop and standardize a reliable and valid Achievement Test in Science. Before developing the test, the researcher reviewed a number of theoretical and empirical literature relevant to the field. For the expert analysis and individual try-out, 14 experts and 14 students were taken. For the purpose of group try-outs, a sample of 100 students was chosen. The test-retest method of reliability was used to determine the reliability of ATS, which yielded a score of 0.87 and the content validity ratio of ATS was 0.99. According to the findings, ATS has a high level of reliability and validity.



## References

1. Aggarwal, V. (2018). Construction and validation of an achievement test for concept assessment in biology. *International Journal of Education*, 9, 32-56. Retrieved 5 March 2021 from <http://ijoe.vidyapublications.com>.
2. Anastasi, A. (1982). *Psychological testing*. 5th Ed., New York: MacMillan Publishing Company.
3. Bala, R. & Singh, G. (2019). Construction and standardization of achievement test in science. *PARIPEX - Indian Journal of Research*, 8(2). 82-84. [www.worldwidejournals.com](http://www.worldwidejournals.com)
4. Bhagat, P. & Baliya, J. N. (2016). Construction and validation of achievement test in science. *International Journal of Science and Research (IJSR)*, 5 (6). 2277 -2280. Retrieved on 5 March 2021 from <http://dx.doi.org/10.21275/v5i6.NOV164808>
5. Chakrabarty, S. N. (2013). Best Split-Half and Maximum Reliability. *IOSR Journal of Research & Method in Education*, 3 (1), 1 - 8. <http://www.iosrjournals.org/iosr-jrme/papers/Vol-3%20Issue-1/A0310108.p>
6. Ebel, R.L. & Frisbie, D.A. (1991). *Essentials of educational measurement*. New Delhi: Prentice, Hall of India Private Limited.
7. Freeman, F.S (1962). *Theory and practice of psychological testing* (3<sup>rd</sup> ed.). New Delhi: Oxford & IBH publishing Co.
8. Knight, J. K. (2010). Biology concept assessment tools: Design and use. *Microbiology Australia*, March, 5 – 8. Retrieved on 5 March 2021 [https://www.colorado.edu/sei/documents/publications/MCDB/biology-concept assessment.pdf](https://www.colorado.edu/sei/documents/publications/MCDB/biology-concept%20assessment.pdf)
9. Kuh, G.D., Kinzie, J., Buckley, J.A., Bridges, B.K., & Hayek, J.C. (2006). What matters to student success: Review of the Literature, mission report for the national symposium on post-secondary student success; spearheading a dialogue on student success Washington, DC: National Secondary Education Cooperative. Retrieved on 13 March 2021 [https://nces.ed.gov/npec/pdf/kuh\\_team\\_report.pdf](https://nces.ed.gov/npec/pdf/kuh_team_report.pdf)
10. Sansanwal, D.N. (2020), *Research methodology and applied statistics*. Delhi: Shipra Publications.
11. Sharma, H.L. & Poonam (2017). Construction and standardization of an achievement test in English Grammar. *International Journal of Advanced Educational Research*, 2(5), 230-235.
12. Sharma, H.L. & Sarita (2018). Construction and standardization of an achievement test in science, 5 (3). 1037 - 1043. Retrieved on 5 March 2021 from <http://ijrar.com>
13. Singh, A.K. (2008). *Tests, measurements and research methods in behavioural sciences*. Patna: Bharati Bhavan Publishers.
14. Singh, A. & Yadav, D. (2018). Construction and standardization of achievement test in Biology. *International Journal of Research in Social Sciences*, 8 (3), p. 18-27. Retrieved from 5 March 2021 from <http://www.ijmra.us>

