

Validation of 8th Class Science Achievement Test in District Sargodha

Muhammad Naqeeb ul Khalil Shaheen*
Nazir Haider Shah**
Rukhsana Sardar***

Abstract

This paper presents evidence for the validation of an Achievement Test of 8th class Science. A teacher-made test was used consisting of 30 multiple choice items from the 8th grade science curriculum. The test was applied to 150 sample students selected by random sampling technique. The test was analyzed by using three techniques for validating the test, these were, individual item's difficulty level technique, technique of discrimination power of the items and distractor analysis technique. The study concluded that twelve items of the test were good and these items can be included in the next test, while, six items of the test were hard and needed to be improved. In addition to this, twelve items of the test needed revision. The study recommended that the teacher should be provided with advance special training in test construction. Furthermore, to make the process of item analysis easy and efficient, a computerized testing system is required to be established.

Keywords: Science Achievement Test, Test Development, Test Validation

* Lecturer, Department of Education, University of Kotli AJ&K.
Email: naqeeb.shaheen@gmail.com

** Assistant Professor, Department of Education, University of Kotli AJ&K.
Email: nazirshah786@gmail.com

*** MS Scholar, Department of Education, International Islamic University Islamabad.
Email: rukhsana.sardar@gmail.com

Introduction

Education is essential for the development of a country. The standard of living of the people of the country depends upon the education system of that country so it is trying its best to raise the standard of education. The basic elements of the whole process of education are the teachers, teaching strategies, curriculum and environment. In the process of educational and personal development, changes occur which must be considered in any meaningful picture of a human being (Shaheen & Kayani, 2015). The tests are used for many purposes in the educational process. The schools use tests as an educational tool to promote individualized instruction. Individualized instruction is the major reason for testing.

The test, if properly conceived can be of assistance in the verification of a pupil's progress. The teacher needs to know about a pupil's achievements to certify these accomplishments to other educational institutions and parents (Balaban & Güneş, 2012). The consistent use of a classroom test provides the teachers with an objective appraisal of each student's learning progress. The test provides the teacher with a student's actual school achievement. The teacher's primary role in the classroom is to teach the students about the development and intellectual growth. To gauge this progress, the teacher must include evaluative techniques (Anastasi & Susana, 2007).

Tests developed by the class teachers to assess the achievements of the students in a particular subject is called a teacher-made test e.g., achievement test in the subject of physics for class 9th. The most instructional relevant tests are those developed by the individual teacher for use with a particular class (Mertens, 2015). A teacher can tailor tests to emphasize the information they consider important and to match the ability levels of their students. If carefully constructed, classroom achievement tests can provide the teacher with accurate and useful information about the knowledge retained by their students.

The standardized test refers to a test that has been expertly constructed, includes expert instructions for uniform administration and scoring. Test plays an important role in evaluation but the validity of the information they provide depends on the care that goes into the development of the tests (Venn, 2000). There are certain steps to construct an achievement test. The first step in constructing an effective achievement test is to identify what are the purposes of the test as they can be used for a variety of purposes. The test may be for placement or diagnostic. It may also be formative or summative. Many aspects of

pupil performance can be measured using a test. All of the intended outcomes of instructions should be considered when planning the test. A list of instructional objectives and specific learning outcomes can be prepared for the test. Table of specification should be prepared according to bloom's taxonomy of objectives (Demir, Gürer, Köksal, & Dolu, 2009).

After the overall objectives of a test are established, the next step in developing a test is to create the detailed test specifications. Test specifications usually include a test description component and a test blueprint component. The learning objectives will determine not only what material to include on the test, but also the specific form the test will take. The test may be objective or essay type or it may include both types of items. In this step, it will be decided that what will be the weightage of different types of items in a test. It involves the development of a test at its initial stage. The items are recorded on cards items prepared according to the table of specifications. A representative sample of all intended outcomes is obtained. Items are prepared to keep in view the proper level of item difficulty.

To validate the 8th class Science achievement test, the research investigated the results of the 8th class science achievement test.

Objective of the Study

The objective of the study was to transform a teacher-made test into a semi-standardized test. For that, a teacher made 8th Class Science Achievement Test was validated.

Literature Review

A teacher uses tests to judge the level of students s/he has attained. Tests are also used to assess the behavior of a student. The dictionary defines the test as 'testum' which is a Latin word from which the word test is derived. In the past, this term was used for the fining pot which was used to melt the metals for refinement. According to Sax (1997), "A test is a task or series of tasks used to obtain systematic observation presumed to be representative of educational or psychological traits or attributes."

The teacher plays an important role in students' intellectual growth and development. He or she is required to use different techniques to measure this progress. Pupils' progress can be verified if the test is taken properly. The regular use of 'tests' in class provides a clear picture of

each student's learning progress. For this, the teachers usually prepare achievement tests for their students.

Achievement tests provide a clear picture of the current status of someone regarding efficiency in a specific area of knowledge or skill. Shaheen, Kayani, and Shah (2015) state that an achievement test is a thing that has been done by someone successfully using his efforts and skill.

Gay (1996) described the classification of achievement tests as oral tests, written tests, prognostic tests, diagnostic tests, power tests, speed tests, accuracy tests, quality tests, range tests, essay type tests, objective type tests, daily tests, weekly tests, monthly tests, term tests, annual tests, arithmetic tests, language tests, reading tests and spelling tests, etc. Based on the method Gay (2009) further classified tests into teacher-made tests and standardized tests which are explained as under.

The teacher-made tests are those which are developed by teachers to check the output of the students in a particular subject or area. e.g. achievement test in the subject of science for 8th class. Tests are taken by the teacher to compare the ability levels of their students and to emphasize the data they consider more important. If tests are carefully constructed and implemented in the class, they provide accurate pictures of students' gained knowledge. Gay (2009) states, "A locally – developed achievement test would likely reflect what was taught to a greater degree than would a standardized test which aims at a wider audience."

A test that has been constructed by an expert and consists of the expert's instructions for consistent evaluation and scoring is known as a standardized test. A standardized test has clear and accurate instructions for administering, scoring and use. Gay (1996) states that "A standardized test is one that is: Measured by specialists and developed according to the subject matter; Field-tested under uniform administration procedures; reconstructed to meet certain criteria and scored and interpreted using uniform procedures and standards."

Characteristics of a Good Test

A test is good when it has certain characteristics these characteristics are validity, reliability, usability and objectivity.

a) Validity

Validity is the degree to which a test measures what it is supposed to measure. A test designed for measuring achievement in the subject of

science will not be valid for measuring the personality, if the test is prepared for 9th class it will not be valid for 10th class. Content Validity, Construct Validity, Concurrent Validity and Face Validity are the most important types of Validity. A good test should be valid in all matters (Shaheen & Kayani, 2015).

b) Reliability

Gay (2000) states that “Reliability is the degree to which a test consistently measures whatever it measures.” Reliability means dependability or trustworthiness. An unreliable test doesn’t provide the same results. Reliability is expressed numerically as a coefficient. A high coefficient value indicates high reliability and a low coefficient value indicates low reliability (Shaheen & Kayani, 2017). If a test is perfectly reliable, the coefficient would be 1.00. Gronlund & Linn (2005) states that the reliability of teacher-made tests lies between 0.60 and 0.80. There are different types of reliability like test-retest reliability, equivalent forms reliability, split-half reliability and rationale equivalence reliability.

c) Objectivity

In the words of Gronlund & Linn (2005): “The objectivity of a test refers to the degree to which equally component scores obtain the same results.” Standardized tests are high in objectivity. In these tests, the resulting scores are not influenced by the scorer’s judgment or opinion. These tests can be accurately scored by scoring machines. For classroom tests constructed by teachers, objectivity plays an important role. In essay type tests, objectivity can be increased by careful phrasing of the questions and by a standard set of scoring rules.

d) Usability

One of the most important characteristics of a good test is usability. It should be constructed such that it can be scored easily. To obtain valid results necessary time must be given during the test. If directions for a test are simple and clear and the time of the test is suitable, it is very easy to administer the test. If directions are unclear and time is short, it badly affects the usability of the test. Gronlund & Linn (2005) state, “In selecting a test and other evaluation instruments, practical considerations cannot be neglected.”

e) Adaptability

Gronlund & Linn (2005) state, “Adaptability means the ability to change, to fit changed circumstances or flexibility”. Adaptability is another feature of a good test. A test should be flexible. It should be

modified according to circumstances and needs. Culture fair tests should be prepared.

f) Economic

A good test should be economic. Sometimes it is given more weight than it deserves. Testing is relatively expensive and cost should not be a major consideration. In large-scale testing programs in which small savings per-pupil add up, using separate answer sheets, machine scoring and reusable booklet reduce the cost appreciably. To select one test instead of another, however, because the test booklets are a few rupees cheaper is a false economy. Validity and reliability are the important characteristics of a test. Gronlund & Lin (2005) state that a test lacking validity and reliability is very expensive at any price.

Item Analysis

In the process of an examination, the item analysis is an important step. The items which are not working well are identified by the use of statistical methods. Sax (1997) defines item analysis as it is the process to find out an item's ambiguity, its level of difficulty and its discrimination index. Once the item analysis information is available, an item review is often conducted.

Methodology

All the students of 8th class of Elementary school who were studying the Science subject the population of the study. A teacher-made test consisting of 30 items each item carries equal marks was administered to the sample of 150 students of the study from 3 schools selected by using a cluster sampling technique. 50 students were selected randomly from each school. Based on their response the process of test validation was completed. Following is the detailed process of validation of the test.

Process of Test Validation

It involves the following steps.

Selection of High and Low Achievers

Firstly, students' mark sheets are arranged in descending order according to the scores they obtained in a test. Secondly, the top 25% of

students are selected who have high scores and they are called high achievers. Similarly, the bottom 25% of students who have low scores are selected and they are called low achievers. Thirdly, the correct responses of both types of achievers on each item are calculated. Sometimes, 27% or 33% of high achievers and low achievers are selected instead of 25%. Ebel and Frisbie (1991) state, “Although upper and lower groups of 27% are best, they are not significantly better than groups of 25% or 33%.” The same criteria were followed in this study.

Item Difficulty

Ebel and Frisibi (1991) define item difficulty as, “Difficulty is defined as the percentage of the group who answered the item correctly.” The difficulty of an item is expressed by its degree of facility or facility index (F). The facility index of an item is defined as the percentage of all candidates making a correct response to particular item difficulty. Facility Index (F) is determined as:

$$F = \frac{NH + NL}{N} \times 100$$

Where

NH = Correct responses of high achievers

NL = Correct responses of low achievers

N = Total no of high and low achievers

If the difficulty index value is high, the item is easy and in case the value is low the item is hard. There are different criteria to accept the item based on item difficulty. Sometimes the range between 20 and 80 percent is considered acceptable. The value between 30% and 70% is also used in most examinations. In the words of Macintosh (1978): “In practice, it has been found that most questions from a variety of examination papers have facility indices covering the range of values from 30% to 70%.” This criterion was followed in this study.

Item Discrimination

Sax (1997) defines item discrimination as, “Discrimination index is the difference between the proportions of individual responding correctly in extreme groups.” Item discrimination is a measure of the extent to

which students who are judged to be good in terms of some standard, answer the item and those who are judged to be poor on the same standard, fail to answer. A discriminating index is to measure the ability of a test item to separate high and low achievers in a test. The index of discrimination was first described by Johnson and Moonan in 1951. It can be determined in terms of formula as:

$$D = \frac{NH - NL}{N/2}$$

Where

NH = Correct responses of high achievers

NL = Correct responses of low achievers

N = Total no of high and low achievers

The discrimination index can range from -1 to +1 it indicates how well an item separates the high and low scores on the total test. Gay (1996) suggests that “A discriminating power of 0.30 is generally accepted as being an adequate value, although the higher is better.” Ebel and Frisbi (1991) have given the following criteria to accept the item based on the discrimination index.

D.I	Interpretation
0.30-1.00	Very good item
0.20-0.29	Reasonably good item but subject to improvement.
0.10-0.19	Marginal items, needing improvement.
-1.00 - .09	Poor item, to be rejected.

This criterion was followed in this study.

Effectiveness of Distracters

A distracter is any of the incorrect answers in a multiple-choice test item. A distracter is to be called a good distracter if it is responded to several students of the lower rank. If this happens, the distracter is retained in the form as it is, but if the distracter is answered correctly by more examinees of the higher group than the examinees of the lower group, the distracter is regarded to be of poor form and the test constructor should re-write the distracter. A distracter must be dropped if it is not answered by any superior or inferior student. Gay (1996) states that a distracter is accepted or rejected under the following criteria.

- A response that is selected as an answer by the number of inferior students than in the superior students.

- If a distractor is not selected by anyone then it is useless and should be replaced.

Data Analysis

The study was aimed at semi standardization of the 8th class Science test. A test was administered consisting of 30 items. The analysis was made based on high achievers and low achievers. The students were ranked and the top 27% were taken in the high achievers' group and the 27% from the bottom were considered the low achievers. The results of the test were analyzed. The following treatments were made for analyzing the test

- Difficulty Level of the individual items
- Discrimination power of the items
- Distractor analysis

Table 1
Difficulty Index of the Items

Item	NH	NL	NH+NL	NH+NL N	D.I.	Status
1	40	21	61	61 / 80	0.763	Acceptable
2	4	13	17	17 / 80	0.213	Hard
3	40	22	62	62 / 80	0.78	Acceptable
4	39	22	61	61 / 80	0.763	Acceptable
5	37	20	57	57 / 80	0.713	Acceptable
6	40	9	49	49 / 80	0.613	Acceptable
7	36	17	53	53 / 80	0.663	Acceptable
8	36	14	50	50 / 80	0.63	Acceptable
9	34	11	45	45 / 80	0.563	Acceptable
10	40	25	65	65 / 80	0.813	Acceptable
11	3	21	24	24 / 80	0.30	Hard
12	5	16	21	21 / 80	0.263	Hard
13	18	18	36	36 / 80	0.45	Acceptable
14	4	17	21	21 / 80	0.263	Hard
15	39	22	61	61 / 80	0.763	Acceptable
16	17	30	47	47 / 80	0.59	Acceptable
17	32	15	47	47 / 80	0.59	Acceptable
18	34	7	41	41 / 80	0.513	Acceptable
19	4	11	15	15 / 80	0.19	Too Hard

20	5	13	18	18 / 80	0.23	Hard
21	21	14	35	35 / 80	0.44	Acceptable
22	34	10	44	44 / 80	0.55	Acceptable
23	38	19	57	57 / 80	0.713	Acceptable
24	8	8	16	16 / 80	0.20	Hard
25	36	7	43	43 / 80	0.54	Acceptable
26	34	15	49	49 / 80	0.613	Acceptable
27	36	24	60	60 / 80	0.75	Acceptable
28	13	13	26	26 / 80	0.33	Acceptable
29	27	8	35	35 / 80	0.44	Acceptable
30	23	18	41	41 / 80	0.513	Acceptable

Table 1 shows that items 2, 11, 12, 14, 19, 20 and 24 were not accepted as they were having the value of Difficulty Index less than or equal to 0.30. Hence, they were considered as hard/ too hard and hence they need to be omitted. The remaining items were acceptable as they were having the value of the Difficulty Index from 0.30 to 1.00.

Table 2
Discrimination Power of the items

Item	NH	NL	NH – NL	NH -NL N/2	D.P	Status
1	40	21	19	19 / 40	0.48	Good
2	4	13	-9	-9 / 40	-0.23	Needs Improvement
3	40	22	18	18 / 40	0.45	Good
4	39	22	17	17 / 40	0.43	Good
5	37	20	17	17 / 40	0.43	Good
6	40	9	31	31 / 40	0.78	Needs Improvement
7	36	17	19	19 / 40	0.48	Good
8	36	14	22	22 / 40	0.55	Good
9	34	11	23	23 / 40	0.58	Good
10	40	25	15	15 / 40	0.38	Good
11	3	21	-19	-19 / 40	-0.48	Needs Improvement
12	5	16	-11	-11 / 40	-0.28	Needs Improvement
13	18	18	0	0 / 40	0	Needs Improvement
14	4	17	-13	-13 / 40	-0.33	Needs Improvement
15	39	22	17	17 / 40	0.43	Good
16	17	30	-13	-13 / 40	-0.33	Needs Improvement

17	32	15	17	17 / 40	0.43	Good
18	34	7	27	27 / 40	0.68	Good
19	4	11	-7	-7 / 40	-0.18	Needs Improvement
20	5	13	-8	-8 / 40	-0.2	Needs Improvement
21	21	14	7	7 / 40	0.18	Needs Improvement
22	34	10	24	24 / 40	0.6	Needs Improvement
23	38	19	19	19 / 40	0.48	Good
24	8	8	0	0 / 40	0	Needs Improvement
25	36	7	9	9 / 40	0.23	Needs Improvement
26	34	15	19	19 / 40	0.48	Good
27	36	24	12	12 / 40	0.3	Needs Improvement
28	13	13	0	0 / 40	0	Needs Improvement
29	27	8	9	9 / 40	0.23	Needs Improvement
30	23	18	5	5 / 40	0.13	Needs Improvement

Table 2 shows that items 1, 3, 4, 5, 7, 8, 9, 10, 15, 17, 18, 23 and 26 were having the value of Discrimination Power between 0.30 to 0.70, hence they were good enough to be accepted. While item 6 had a value of Discrimination Power more than 0.70 and the remaining items were having a value less than 0.30 hence they needed improvement.

Table 3
Distracter Analysis

Item	Responses of HA				Responses of LA				Correct Choice
	A	B	C	D	A	B	C	D	
1	0	0	0	40	11	0	8	21	D
2	4	29	6	1	13	11	15	1	A
3	0	40	0	0	5	22	6	7	B
4	1	0	0	39	8	5	5	22	D
5	3	0	37	0	5	10	20	5	C
6	0	0	40	0	7	16	9	8	C
7	0	38	0	2	7	17	11	5	B
8	0	1	3	36	10	6	10	14	D
9	4	2	34	0	6	11	11	12	C
10	0	0	0	40	0	15	0	25	D
11	36	2	0	4	21	3	9	7	A
12	2	0	5	33	6	17	16	1	C
13	4	18	0	18	1	11	10	18	D
14	34	0	4	2	11	9	17	3	C
15	0	0	39	1	3	3	22	12	C
16	17	2	7	4	30	10	0	0	A
17	3	5	32	0	9	12	15	4	C
18	4	0	2	34	17	7	9	7	D

19	2	4	29	5	7	11	17	5	B
20	0	4	31	5	5	18	13	4	C
21	0	21	6	13	8	14	7	11	B
22	0	34	6	0	5	10	21	4	B
23	0	1	38	1	2	12	19	7	C
24	5	9	18	8	8	20	4	8	D
25	36	4	0	0	7	9	10	14	A
26	6	34	0	0	16	15	9	0	B
27	2	36	0	2	0	24	15	1	B
28	0	11	13	16	5	12	13	10	C
29	5	8	0	27	5	19	8	8	D
30	0	14	23	3	4	12	18	6	C

Table 3 shows the results of distractor analysis. This indicated that items # 3, 4, 5, 6, 7, 21 and 25 used good distracters hence in these items no change of distractor is required. On the other hand for items # 1, 2, 8, 9, 11, 12, 13, 14, 17, 23, 24, 26, 28 and 30 did not use appropriate distractors in the form of options B, D, B, A, B, D, A, D, A, C, D and A respectively, hence the mentioned options require to be replaced by good distractors.

For item 10 options A and C are not good distractors. Similarly, for item 15 options A and B, for item 16 options C and D, for item 18 options B and C, for item 19 options A and D, for item 20 options A and D, for item 22 options A and D, for item 27 options A and D and item 29 options A and C are not good distractors as a fewer number of students choose them. Hence, all these distractors also require replacement.

Conclusions

It was concluded that:

1. Items 1, 3, 4, 5, 6, 7, 8, 9, 10, 17, 18 and 26 were good and these items can be included in the next test.
2. Items 2, 11, 12, 14, 19 and 24 were hard and needed to be improved.
3. Items 13, 15, 16, 20, 21, 22, 23, 25, 27, 28, 29 and 30 should be revised and improved.

Recommendations

The following recommendations are made based on the findings and conclusions of the study.

- The teacher should be provided with advance special training in test construction. So that they may be able to construct objective type test and be able to use new approaches to the assessment and evaluation.
- To make the process of item analysis easy and efficient, a computerized testing system should be established.
- The items in the final test having discrimination indices below 0.20 should be improved to be used in a subsequent administration.

References

- Anastasi, A. & Susana, U. (2007). *Psychological testing*. Delhi: Dorling Kindersley (Pvt.) Ltd.
- Balaban, M. & Güneş, M. H. (2012). Portfolio assessment in cooperation with teachers, students and parents in a science and technology course. *Eurasian Journal of Educational Research*, 49, 289-310.
- Demir, S., Güreç, C., Köksal, T. & Dolu, O. (2009). Kavram Oluşturma ve Ölçüm. (Ed.) Kaan Böke. *Sosyal Bilimlerde Araştırma Yöntemleri*. İstanbul: Alfa Yayıncılık.
- Ebel, L. and Frisbie, A. (1991). *Essential of educational measurement*. New Delhi: Prentice Hall of India (Pvt.) Ltd.
- Gay, L. R. (1996). *Educational evaluation and measurement*. New York: McMillan Publishing Company.
- Gay, L. R. (2000). *Educational research*. Islamabad: National Book Foundation.
- Gay, L. R. (2009). *Educational research*. Islamabad: National Book Foundation.
- Gronlund, N. & Linn, R. (2005). *Measurement and assessment in teaching*. Delhi: Pearson Education (Pvt.) Ltd.
- Johnson, M. O. & Moonan, W. J. (1951). Recent developments in statistical theory. *Review of Educational Research*, 21 (5), 389-414.
- Macintosh, H. G. (1978). *Techniques and problems of assessment*. London: Edward Arnold Ltd.
- Mertens, D. M. (2015). *Research and evaluation in education and psychology: integrating diversity* (4th ed.). Sage Publications.
- Sax, G. (1997). *Principles of educational and psychological measurement and evaluation*. Washington: International Thomson Publishing Company.
- Shaheen, M. N. K. & Kayani, M. M. (2017). improving students' attitudes towards biology as a school subject: do the instructional models really work? *Journal of Applied Environmental and Biological Sciences (JAEBS)*, 7(1), 170-179.

Shaheen, M. N. K. & Kayani, M. M. (2015). Improving students' achievement in biology using 7e instructional model: An Experimental Study. *Mediterranean Journal of Social Sciences*, 4 (S3), 471-481. doi: 10.5901/mjss.2015.v6n4s3p471

Shaheen, M. N. K., Kayani, & M. M., Shah, N. H. (2015). Teaching of science at secondary level: an analysis of teachers' classroom practices. *International Journal of Innovation in Teaching and Learning (IJITL)*, 1(1), 25-34.

Venn, J. (2000). *Assessing students with special needs*. London: Prentice Hall International.

Citation of this Article:

Shaheen, M. N. K., Shah, N. H., & Sardar, R. (2020). Validation of 8th class science achievement test in district Sargodha, *Journal of Science Education*, 2(1), 63-77.