

Achievement Test

Anjali Arora Soni, Jasdeep Kaur

Assistant Professors, D.D. Jain College of Education, India

Abstract

Assessment data is used on routine basis for making educational decisions about the students. Purposes of assessment are decision making, screening, placement, programme planning, evaluation, individual progress monitoring and many more. Thus, the present study was conducted to construct an achievement test (teacher made test) in commerce. The test was developed on four selected topics viz.- Nature and Purpose of Business, Social Responsibility of Business, Sources of Business Finance and Forms of Business. The need was felt to develop this test to evaluate the outcomes of instructional strategy. The test was made to measure knowledge, comprehension and application of the concepts of Commerce students preparing for Competitive Exams.

KEYWORDS: Business, Social Responsibility of Business, Sources of Business Finance and Forms of Business

The achievement tests are controlling process for teaching learning activities. These help in evaluating the effectiveness of teaching instructions. Achievement tests are past and present oriented which register the degree of learning or achievement after instructions. According to *Dowine and Heath (1974)* any test that measures the attainments or accomplishments of an individual after a period of learning is called achievement test. It can also be thought of as a sample of indicator of a student's knowledge taken at a particular point in time.

Therefore, achievement tests propose to measure what and how much pupils have learnt as a result of formal and informal instructions. These are utilized as evaluating courses of study or efficiency of teacher and teaching methods or other educational factors. Achievement tests differ from intelligence or aptitude tests in that the former measures the quantity or quality of learning attained in a subject of study or group of subjects after a certain period of instruction and the latter measure pupil's innate capacity for attainment or accomplishment independent of learning.

Lindeman (1967) classified achievement tests in following categories:

- (i) Teacher Made Test
- (ii) Standardized Test

Teacher made tests are frequently the basis of evaluating the students' progress in school. These are more specifically focused and they usually reflect the content of a particular unit or course. The teacher made test is tailored to measure the achievement of students and intended objectives for them after completing a series of learning tasks for the course. According to *Anastasi (1988)* "undoubtedly the largest number of tests covering the content of specific courses is prepared by the instructors for use in their own

classroom. Most schools focus on student progress as the ultimate criterion therefore, it is important to evaluate the status of pupils expertly.” It was indicated by *Anastasi (1988)* that the preparation of local classroom tests can substantially be improved by an application of the techniques and accumulated experience of professional teachers. It is necessary that teachers in the preparation of their own classroom test follow specific guidelines to ensure quality test output and attaining the goal of measuring effectively what students have learnt.

On the other hand, standardized tests are particular kinds of tests, different from final examination a high school teacher might design for his/her course. When talking about tests, “standardized” simply means that everyone who takes the test is given the same amount of time and sees the same or very similar test questions. “Standardized” also means scoring is done very carefully so that the test scores do not depend upon who happens to be doing the scoring. These tests are widely used because, by and large, they have shown to be an efficient way to collect information about what people know and can do. These are objective in the sense that test scores do not depend to any great extent on who happens to score the answers. These are valid in that these often provide relevant and useful data for making decisions about mastery of a body of knowledge and skills and potential for success.

Good (1959) defined standardized test as “a test for which content has been selected and checked empirically, for which norms have been established for which uniform methods of administering and scoring have been developed and which may be scored with a relatively high degree of objectivity.”

Therefore a standardized test is a test administered and scored in a consistent manner. The tests are designed in such a way that the “questions, conditions for administering, scoring procedures, and interpretations are consistent” and are “administered and scored in a predetermined, standard manner.”

DEVELOPMENT OF ACHIEVEMENT TEST IN COMMERCE

For the present study the investigator could not lay hands on appropriate standardized achievement test in commerce on the selected topics. The need was felt to develop one so as to evaluate the outcomes of instructional strategy. The test was developed on four selected topics viz.-

- (I) Nature and Purpose of Business
- (II) Social Responsibility of Business
- (III) Sources of Business Finance
- (IV) Forms of Business

1.1 PLANNING OF THE TEST

It was the first step of test construction. The test construction specified the broad and specific objective of the test in clear terms. Nature of the questions was given weight age on each and every component of the subject matter as far the curriculum of a particular class was concerned. *Stanley and Hopkins (1990)* observed that the planning stage of a test should include nature of the test, test items and a statement of conditions under which it will be administered.

On the basis of planning strategies and nature of problem following preliminary conditions were set:-

1.1.1 Test Purpose

The purpose of test was to measure the acquisition of concepts of Commerce. The test was made to measure knowledge, comprehension and application of the concepts of Commerce students preparing for Competitive Exams.

1.1.2 Content of the Test

The test devised covered the contents from the following four topics undertaken for the study:

- (I) Nature and Purpose of Business
- (II) Social Responsibility of Business
- (III) Sources of Business Finance
- (IV) Forms of Business

1.1.3 Types of Test Items

All the items were of multiple choices. Each item has four alternatives out of which one is correct.

1.1.4 Blue Print of the Test

After deciding about the objectives of the test a pre-assessment had been made of the items to be put in the test. Nature of questions was given weight age on each and every component of subject matter as far as the curriculum is concerned.

After studying thoroughly the syllabus prescribed for competitive exams, the main content area was identified and the major course instructional behavioral objectives were specified. Adequate provisions were made for evaluating all the important outcomes of instruction. A tentative blue print was prepared. It was then examined by two Atam Devki Niketan school teachers, two other experts (dealing with subject) of Ludhiana and three experts from the Colleges of Education of Ludhiana, for giving their opinion about the items and weightage given to each content area. For the present test, a blue print with three dimensions i.e. knowledge of content, comprehension and application was prepared as under:

Table: Blue Print of Achievement Test (First Draft)

S.No	Topic	Weightage			Total no. of items
		Knowledge	Comprehension	Application	
1.	(I) Nature & Purpose of Business	3	4	2	9
2.	(II) Social Responsibility of Business	4	5	3	12

3.	(III) Sources of Finance	6	4	1	11
4.	(IV) Forms of Business	4	1	2	7
	Total	17	14	8	39

1.2 FIRST DRAFT OF THE TEST

Keeping the blue print for guidance, items were written in English language to cover the designated content. Many more items were written initially than would be included in the final form of the test because many items would be omitted in the process of item analysis. While writing the items following points were kept in mind:

- The items were phrased in such a manner that there was no ambiguity regarding their meaning.
- The items were not concerned with the trivial aspect of the subject matter, that is, it only measures the significant aspects of knowledge and understanding.
- Interdependence among the items was avoided.
- Two objectives were not put in the same item.
- The keyed responses were randomly varied in position.
- As far as possible the items did not encourage guess work.

The items written were reviewed, criticized and revised after thorough discussion with experts for the content validity. Most of the revisions involved eliminating ambiguities, clarifying wordings and strengthening weak alternatives. The distracters of multiple choice items were distributed equally such that these should also be in a position to discriminate between students. The preliminary achievement test consisted of 39 items, a few easy and a few difficult ones.

Scoring

A scheme of evaluation was prepared for scoring the papers. One point of credit was given for each correct response.

Preliminary try out of the Test

The test was administered to thirty students of Atam Devki Niketan. During data collection, proper rapport was established and maintained with the students. The respondents were explained how to take the test. Every reasonable precaution was taken to ensure normal conditions during the administration of the test.

After preliminary try out, problems faced by the students were noted and given the due consideration at the time of revision of first draft. On the basis of the performance of the students, discussions were held with the teachers and students individually.

Some irrelevancies and ambiguities which were wholly unsuspected by the investigator before preliminary try out, came to light after it. In the light of the view of Commerce teachers, the achievement test was reviewed properly. As a result of the discussion 9

items were clearly discarded and in 5 items slight changes were made to improve clarity in the wording.

Table: Serial Number of Dropped Items of the Achievement Test

S.No.	Topic	Serial No. of Dropped Items	No. of Items Left
1.	(I) Nature & Purpose of Business	1, 2 & 5	6
2.	(II) Social Responsibility of Business	11 & 16	10
3.	(III) Sources of Finance	24, 28 & 31	8
4.	(IV) Forms of Business	38	6
	Total		30

1.3 ITEM ANALYSIS

In item analysis suitability of each item had been seen statistically one by one. Item analysis is a set of procedures that is applied to know the indices for the truthfulness of items. In other words, item analysis is a technique through which those items which are valid and suited to the purpose are selected and the rest are either eliminated or modified to suit the purpose. In brief, it can be said that item analysis demonstrate how effectively a given test item functions within the total test. The main objectives of item analysis are enumerated below:

- Item analysis indicates which items are difficult, easy, moderately difficult or moderately easy. In other words, it provides an index of the difficulty value of each item.
- It also provides indices of the ability of the item to discriminate between high and low. In other words, item analysis indicates the discrimination value of each item. This is known as discrimination index.

For item analysis a sample of 30 students from Atam Devki Niketan , Ludhiana was selected. The test was administered to them. The aim of this try out was to provide data for item analysis. Clear and detailed instructions were given for proper administration of the test.

For the present test 27% upper and 27% lower cases were considered for calculating difficulty value and discrimination value of each item. *Kelley (1939)* showed that the product moment correlation between a test item score and the total score could be estimated by using only the tails of the distribution and he also showed that the most efficient division to use was the top and bottom 27 percent tails. This suggestion was

followed. The valued test papers were arranged from the highest score to the lowest score. The top 27% of the students were kept in one group i.e. the high group. The poorest 27% of the students were kept on the other group i.e. the low group. The middle group consisting of 46% of papers was kept aside since the two extreme groups high and low only were needed for item analysis.

For calculating the difficulty value (D.V.) and discriminating power (D.P.) of items following formulae were used:

$$D.V. = \frac{RU + RL}{N} \quad D.P. = \frac{RU - RL}{N/2}$$

Where

RU = Total Number of Right Responses in the Upper Group

RL = Total Number of Right Responses in the Lower Group

N = Total Number of Students in Both the Groups (28)

For the value of D.V. and D.P. for each item of the Achievement Test kindly refer to the following table:

ITEM	UPPER 27%	LOWER 27%	DIFFERENCE	DISCRIMINATION INDEX	DIFFICULTY VALUE
1	8	7	1	0.125	0.9075
2	7	6	1	0.125	0.8125
3	7	4	3	0.375	0.6875
4	6	2	4	0.50	0.50
5	5	2	3	0.125	0.4375
6	7	4	3	0.375	0.6875
7	6	4	2	0.25	0.625
8	5	3	2	0.25	0.500
9	7	5	2	0.25	0.75
10	4	1	3	0.375	0.3125
11	3	0	3	0.375	0.1875
12	6	2	4	0.50	0.50
13	7	4	3	0.375	0.6875
14	5	2	3	0.375	0.4375

15	8	3	5	0.625	0.6875
16	8	8	0	0	1
17	3	1	2	0.25	0.25
18	5	1	4	0.50	0.375
19	5	3	2	0.25	0.50
20	5	3	2	0.25	0.50
21	8	3	5	0.625	0.6875
22	3	1	2	0.25	0.25
23	6	2	4	0.50	0.50
24	8	7	1	0.125	0.9075
25	5	3	2	0.25	0.50
26	3	1	2	0.25	0.25
27	7	4	3	0.375	0.6875
28	7	6	1	0.125	0.8125
29	8	3	5	0.625	0.6875
30	5	1	4	0.50	0.375
31	7	6	1	0.125	0.8125
32	6	2	4	0.50	0.50
33	5	2	3	0.375	0.6875
34	7	5	2	0.25	0.75
35	5	3	2	0.25	0.50
36	6	3	3	0.375	0.5625
37	7	4	3	0.375	0.3125
38	0	0	0	0	0
39	4	1	3	0.375	0.3125

1.4 FINAL DRAFT OF THE TEST

Final draft of the achievement test was prepared on the basis of item analysis for difficulty value and discriminating power. According to *Garret (1981)* "As a general rule, validity indices of .20 or more are regarded as satisfactory." Following this suggestion, the items with difficulty value from .25 to .75 and discriminating power from .20 to .90 were retained in the final draft of the achievement test. On the basis of this criterion the items at the serial number 1, 2, 5, 11, 16, 24, 28, 31 & 38 were rejected from the achievement test and the remaining 30 items were retained

Table: Distribution of Items of Achievement Test (Final Draft)

S.No	Topic	Weightage			Total no. of items
		Knowledge	Comprehension	Application	
1.	(I) Nature & Purpose of Business	2	2	2	6
2.	(II) Social Responsibility of Business	4	4	2	10
3.	(III) Sources of Finance	5	2	1	8
4.	(IV) Forms of Business	3	1	2	6
	Total	14	9	7	30

1.5 RELIABILITY

Reliability refers to the consistency of scores or measurement which is reflected in the reproducibility of the scores. A test is said to be consistent over a given period of time when all the examinees retain their same relative ranks of two separate testing with the same test. In other words, reliability tests that to what extent individual differences of scores can be assigned to chance error.

In the words of *Anastasi (1951)*, "Reliability refers to the consistency of scores obtained by the same individuals when re-examined with the same test on different occasions or with different sets of equivalent items or under other variable examining conditions." *Ebel (1979)* defines it as the consistency which a set of test scores measure whatever they do measure.

There are four procedures in common use for computing reliability. These are:-

1. Alternative or Parallel Form Method
2. Split-half Method
3. Rational Equivalence Method

4. Test-Retest Method

1. Alternate or Parallel Form Method

Parallel form reliability requires the administration of parallel or equivalent forms of a test to the group, and the calculation of the correlation between the two sets of scores. Often both forms are administered in one sitting or in two sessions separated by a very short time period. *Guliksen (1974)* speaks of parallel tests in terms of statistics. In his opinion, "Parallel tests have equal means, equal variances and equal inter-correlations with another." In the opinion of *Guilford (1956)*, "The alternate form method indicates both equivalence of content and stability of performance." This method is one of the appropriate methods of determining the reliability of educational and psychological tests. But there are some practical difficulties in this method. It is difficult to have two parallel forms of a test. In certain situations it is almost impossible.

2. Split-half Method

The split method provides an estimate of the extent to which a test is internally consistent and permits a reliability coefficient to be obtained from a single administration of the test. This method involves the splitting of a single test into two comparable sets of items and calculating the correlation between them. A common practice is to allocate odd-numbered test items to one set and even-numbered items to the other set. The advantage of the split method is that all data necessary for computation of reliability coefficient are obtained in single administration of the test. Thus the variability produced by the differences in the two administrations of the same test is automatically eliminated. Thus, a quick estimate of reliability is made. That is why; *Guilford* and *Fruchter (1973)* have described it as on-the-spot reliability. The disadvantage of this method is that it is not appropriate where speed is an important factor in the test, since it is assumed that most or all individuals being tested have had a chance to attempt all test items. One of the problems in this approach is that because of variations in consistency, different values of the correlation coefficient may be obtained as a result of splitting the test in different ways.

3. Rational Equivalence Method

This method is also known as 'Inter-Item Consistency'. It is a method based on single administration. Two forms of a test are said to be equivalent when corresponding items are inter-changeable and when item-item correlations are the same in both the forms. The inter item consistency is influenced by two sources of variance; (1) content sampling, and (2) heterogeneity of the behavior domain sampled. The more homogeneous the domain, the higher is the inter-item consistency. *Guilford (1956)* maintains that it is based on the assumption of unifactor test and parallel items and in the absence of such a test the use of the formula is well within doubts.

4. Test-Retest Method

In test-retest reliability the single form of the test is administered twice on the same sample with a reasonable time gap. In this way, two administrations of the same test yield two independent sets of scores. The two sets, when correlated, give the value of the reliability coefficient. The reliability coefficient thus obtained is also known as the temporal stability coefficient and indicates as to what extent the examinees retain their relative position as measured in terms of the test scores over a given period of time.

Guilford (1956) writes in this regard that, “A retest coefficient of correlation tells us nothing concerning the internal consistency of a test. The key concept for this procedure is that of stability. It answers the question concerning how stable or dependable are the measurements over a period of time. **Morley (1970)** remarked, “The test-retest method is the only feasible approach to the establishment of the reliability of the test. The answer to one question given by a respondent in two instances can be compared for estimating consistency.”

Keeping the above views **test-retest method was found better option** as assumptions of unifactor test and parallel items were not met so option of *Kunder-Richardson* formula was dropped. Then again as one form of the test was constructed, parallel form method was ignored. Lastly, the split half method was discarded as items were not arranged on the basis of item type.

The test was administered to 26 Commerce students of Ludhiana those are preparing for competitive exams who were not included in the experimental sample of population. With a gap of 15 days the test was again administered to the same group.

1.6 VALIDITY

The degree of validity is the single most important aspect of a test. Validity can be best defined as the degree to which a test is capable of achieving certain aims. The validity of the test is determined by measuring the extent to which it matches with a given criterion. According to **Anastasi (1951)**, “The question of test validity concerns what the test measures and how well it does so.” The process of validity involves checking the agreement between the responses elicited by each question item against the criterion. Validity of a test can be established in five ways:

1. Content Validity

Content validity also designated by other terms such as intrinsic validity, relevance, circular validity and representatives. When a test is constructed so that its content measures the same objectives decided for the whole test, the test is said to have content validity. Thus, the content validity is concerned with the relevance of the contents of the items, individually and as a whole. **Anastasi (1988)** has said that content validity, “involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured.” Content validity is needed in the tests, which are constructed to measure how well the examinee has mastered the specific skills or a certain course of study. Content validity of a test is examined in two ways:

- (i) By expert’s judgments and
- (ii) By statistical analysis.

In first case, if the investigator wants to examine the content validity of a test, the items of the test will be submitted to a group of subject-matter experts. These experts will judge whether or not the items represent all the important area of the content to be measured, whether or not some additional items should be added for complete coverage, what should be relative weights of the items of a particular area etc. The validity of the items will depend upon consensus judgments of the majority of the subject-matter experts.

Statistical methods may also be applied to ensure that all items measure the same thing i.e. a statistical test of internal consistency may provide evidence for the content validity. Another statistical technique for ensuring content validity may be to correlate the scores on the two independent tests, both of which are used to measure the same thing.

Content validity is most appropriately applied to achievement tests or the proficiency test.

2. Construct Validity

Anastasi (1988) has defined it as the extent to which the test may be said to measure a theoretical construct or trait. A construct is a sort of concept which is formally proposed with definition and is related to empirical data. According to *Nunnally (1967)*, a construct indicates a hypothesis which tells us that a variety of behaviors will correlate with another in studies of individual differences and or will be similarly affected by experimental treatments.

3. Criterion-related Validity

Criterion-related validity is a very common and popular type of test validity. As its name implies, criterion-related validity is one which is obtained by correlating the test scores with scores obtained on a criterion available at present or to be available in the future. The criterion is defined as an external and independent measure of essentially the same variable that the test claims to. In this sense, by way of defining the validity of a test, *Cureton (1965)* has said that the validity of a test is an estimate of the correlation coefficient between the test scores and the criterion scores. There are two sub types of criterion-related validity:

Predictive Validity

In predictive validity a test correlated against the criterion to be made available sometime in the future. The predictive validity coefficient is a Pearson Product-Moment correlation between the scores on the test and an appropriate criterion where the criterion measure is obtained after the desired lapse of time.

Concurrent Validity

Concurrent validity is very similar to predictive validity except that there is no time gap obtaining test scores and criterion scores. The test is correlated with a criterion, which available at the present time.

Regarding establishing the validity of a test *Morley (1970)* remarked, "At the most elementary level, it is necessary for all the tests to have content validity i.e. each question must be related to its topic. There must be an adequate coverage of the overall topic. The questions must be cleared and unambiguous." Thus, content validity is concerned with the relevance of the contents of the items, individually and as a whole. Each individual item or content of the test should correctly and adequately measure the trait or the variable in question and the test as a whole should contain only the representative items of the variable to be measured by the test.

Thorndike and Hagen (1970) stated that, "The problem of content validity is parallel to the problem of preparing a blue print for a test and then building a test to

match blue print.” In the present study, the questionnaire responses of the test were validated against the competencies. Content validity was found by relating each item carefully against the competencies. To confirm validity, the test items along with a list of competencies to be developed in the learners were given to a panel consisting of 10 experts in the subject matter and three in test items. Out of ten subject experts, 6 experts also solved the test so that the scoring key could be verified. Only those items were kept in the test draft for which there was complete agreement among the experts. The content validity was ascertained by preparing a blue print of the test items indicating the weightage given for the objectives and competencies by these items.

1.7 TIME LIMIT FOR THE TEST

When the test was given to an equivalent group for item analysis the time taken by each student was noted to have an estimate of the time required for the completion of the test. The mean of the time of first ninety percent students was calculated which came out to be 35 minutes. Then the test was again given to eighty students of equivalent group. The students were divided in five groups and were given time of 40 minutes, 45 minutes, 50 minutes, 55 minutes and 60 minutes to complete the test. Then standard deviation of each group was calculated. The standard deviation of the group, who have completed test in 30 minutes was maximum. Hence the time limit for the completion of the test was set at 30 minutes.

Reference

Aggarwal (2008). *Teaching of Commerce: A Practical Approach* (2nd ed). UP: Vikas Publishing House Pvt. Ltd.

Gupta, U. C. (2007). *Teaching of Commerce*. New Delhi: Khel sahitya Kendra.

Kumar, K.L. (1997). *Educational Technology*. New age International Pvt. Ltd., Mumbai.

Mohan, R. (2010 ed) “*Research methods in education*” (Revised ed.). Neelkamal Publications Pvt Ltd, New Delhi.

Rao, B.; Digumarti & Latha, D.P. (1995). *Achievement in English*, New Delhi, Discovery Publishing House.

Rao, S. (2007). *Teaching of Commerce*. New Delhi: Anmol Publication.

Siddiqui, M.H. (2009). *Techniques of Classroom Teaching*. New Delhi: APH Publishing.

Singh, Y. K. (2005). *Teaching of commerce*. New Delhi: A.P.H. Publishing Corporation.

Singh, Y. K. (2009). *Teaching of Commerce*. New Delhi: APH Publishing Corporation.

Webliography

<http://kailash-kandpal.blogspot.in/2011/11/achievement-test-business-studies.html>.

<http://www.ericdigests.org/pre-9213/classroom.htm>