CONSTRUCTION AND EVALUATION OF RELIABILITY AND VALIDITY OF REASONING ABILITY TEST

Mehraj A. Bhat

Department of Education, Aligarh Muslim University, U.P., India.

A B S T R A C T

This paper is based on the construction and evaluation of reliability and validity of reasoning ability test at secondary school students. In this paper an attempt was made to evaluate validity, reliability and to determine the appropriate standards to interpret the results of reasoning ability test. The test includes 45 items to measure six types of reasoning in one hundred and ten secondary school students. Content validity was evaluated by more than 25 experts and validity was calculated by correlation between the score of each dimension and total score of the test. To know the discrimination validity for each dimension of the test, 't' test for two independent samples was used (high group and low group). The reliability of the test was tested by calculating Alpha Cronbach. To identify those students who are competent or incompetent in the reasoning ability percentiles were used to determine the adequate cutoff score for the test. Overall it is concluded that the test has good construct and discrimination validity. Moreover, all the values of reliability coefficient for each dimension are highly significant.

Keywords: Construction, evaluation, reliability, validity, reasoning ability.

INTRODUCTION

Since the evolution of human beings, reasoning ability has been used as an important element to solve their day today problems. It has been recognized as the core element of human nature. Its expression can be found in the teaching of Socrates, Confucius and others (Chen, 2000). The goal of education is to equip its citizens with the ability to reason out. Therefore the development of reasoning skills, its improvement and various approaches have brought out immediate concerns of educators, psychologists, and philosophers for decades (Kemler, 1998).

Reasoning occupies an important place in our daily life. We take up its help consciously or unconsciously every day. All our activities and decisions are based on our reasoning. An individual is guided in taking a decision only after he reasons out the matter in his mind (Fatima 2008). This is because almost everything we do and think involves drawing conclusions. When we learn, criticize, judge, infer, evaluate, optimize, apply, discover, imagine,

* Corresponding Author:
Email: mehrajsc@gmail.com

devise, and create, we draw conclusions from information and form our beliefs (Leighton, 2004). In today's complex world, the ability to think and reason logically is essential for everybody. The ability to reason is indispensable when problem solving skills are required. Thus in situations in which experienced operations and algorithms for problem solution are not available or cannot be retrieved. Without reasoning, already acquired knowledge and experiences could not be applied to new situations.

Our understanding of vital social economic and political problems of today is largely dependent upon reasoning and mankind struggle against poverty, ignorance and diseases, against war, racial prejudice and cruelty is being carried on through powerful reasoning. It is a powerful source of individual efficiency and wellbeing. It is through reasoning that the individual is able to rise above life of impulse and raw emotion, to predict the effects of his course of action and to plan his conduct for personal and social benefit.

Reasoning skills are recognized as the key abilities for human being to create, learn, and exploit knowledge. These skills are also an important factor in the process of

human civilization. The significance of reasoning skills has been of great concern in educational settings and in the world of work. It becomes increasingly important to improve reasoning ability through lifelong learning in response to such challenges and lead a meaningful life, and construct a rational better world (Shu, 2000). Therefore, current educational systems across the world have recognized the need to enhance students' reasoning ability (Wu, 2001). It plays a significant role in one's adjustment to one's environment. It is essentially a cognitive ability and is like thinking in many aspects (Bandhana, 2012). Reasoning can be categorized as (1) Inductive reasoning (2) Deductive reasoning (3) Analogical reasoning (4) Linear reasoning (5) Conditional reasoning (6) Abduction as reasoning (7) Syllogistic reasoning (8) Pros-vs.-cons reasoning (9) Set-based reasoning(10) Systematic reasoning (11) Cause and effect reasoning (12) Comparative reasoning (13) Decompositional reasoning and (14) Analytical reasoning. Cavallo (1996) found that reasoning ability best predicted students' achievement in solving genetic problems. The study carried out by Lawson and Thompson indicated that misconceptions are consistent and significantly related to the reasoning ability. Moreover, the students with the highest level of formal reasoning might change their alternative conception more easily (Lawson, 1998). (Yenilmez, 2006) investigated the effect of gender and grade level on students' reasoning abilities. Results showed that boys have higher scores than girls on proportional, probabilistic and combinational reasoning, whereas girls have higher scores on controlling variables and correlation reasoning. It was also found that there was a statistically significant gender difference in favour of boys for proportional reasoning.

Objectives:

To construct reasoning ability test secondary school students.

To evaluate the validity of reasoning ability test.

To evaluate the reliability of reasoning ability test.

To determine the appropriate standards to interpret the results of reasoning ability test.

METHODOLOGY

The method adopted for the present study can be categorized as descriptive statistical in nature. Descriptive research describes and interprets the current status, it is concerned with conditions or relationship that exist, practices, that prevail, beliefs, points of view or attitudes that are held, processes that are going on, effects that are being felt or trend that are developing. The process of description as employed in this research study goes beyond mere gathering and tabulation of data. It involves an element of interpretation of the meaning or significance of what is described. Thus, description is combined with comparison or contrast involving measurement, classification, interpretation and evaluation.

Sample: The samples of the study comprised of 110 secondary school students currently enrolled in class 10th of different (Govt./Private) schools of South Kashmir of Jammu and Kashmir. This study was delimited to students of class 10th. Secondly the age range of the members of the population is 15-16 years.

Stages of tool construction: As with the test classification, there is no total agreement of experts about the precise steps for test construction. Nevertheless, when constructing a test, it is necessary to go through a number of stages in order to ensure its good quality (Alderson, 1995). Although their needs a proper procedure for test construction. The graphical representation for the stages of tool construction as depicts in figure 1.
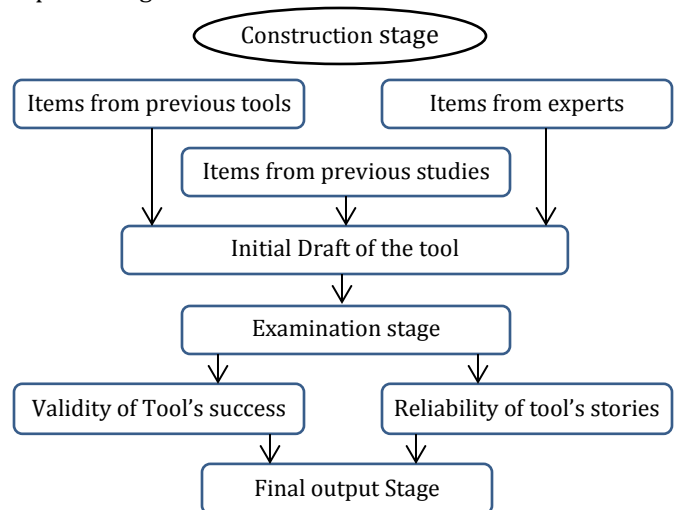

Fig 1. Stages of Tool Construction.

Preparation of preliminary draft: Once defining the reasoning ability and its types, the items associated to six dimensions were selected. Each item was selected according to the nature of the dimension. For the selection of the items different books related with the reasoning were used (jøsang, 2008; Jeotee, 2012; Aggarwal, 2013), besides that the researcher used previous tools and studies related with reasoning and also the researcher obtained assistance from many experts in

education and psychology about the items which help to measure it. While selecting items, the nature of item measured the desired dimension of reasoning were taken into consideration. In this way the initial draft was prepared and 72 items were included in the scale. Then, draft items were given to experts from different universities who were well versed in the field and scale construction with a request to review the statements and evaluate their content accuracy coverage, editorial quality and suggestions for additions, deletion and modification of items. Based on 80% unanimity of the experts, 45 items were included in the format of the scale (Table 1).

Table 1. Distribution of dimensions with respect to their Items.

Sr. No.	Dimensions	No. of Items
1.	Analogical Reasoning	9
2.	Linear Reasoning	5
3.	Conditional Reasoning	8
4.	Deductive Reasoning	9
5.	Inductive Reasoning	7
6.	Cause and Effect Reasoning	7
	Total	45

Item analysis: The initial format with 45 items on a four alternative responses was administered on the sample, each question carried one point (1) for right answer and zero (0) point for wrong answer. The response sheets received from the students were arranged from maximum on the basis of overall score. The obtained data were used to assess the difficulty level and discriminating power of test items.

Difficulty level: To calculate the difficulty level the researchers used the following formulae:

Difficulty equation: $D = \frac{p1+p2}{n}$

p1= the number of students who give right answer in high group, p2= the number of students who give right answer in low group, n= the total no. of students of high group

and low group. The results of the said test with respect to difficulty level of items are given below.

Table 2. Item's difficulty level.

Item No.	Difficulty Value	Item No.	Difficulty value
1	0.83	24	0.73
2	0.57	25	0.87
3	0.80	26	0.77
4	0.77	27	0.87
5	0.70	28	0.10
6	0.80	29	0.67
7	0.30	30	0.97
8	0.67	31	0.83
9	0.73	32	0.63
10	0.87	33	0.70
11	0.73	34	0.70
12	0.33	34	0.67
13	0.63	36	0.63
14	0.53	37	0.43
15	0.97	38	0.83
16	0.50	39	0.47
17	0.63	40	0.63
18	0.87	41	0.40
19	0.40	42	0.50
20	0.77	43	0.50
21	0.37	44	0.43
22	0.63	45	0.23

General guidelines for difficulty value: Low value of difficult index means that the item is a very difficult one, e.g., if D.V = 0.20 it means that only 20% answered correctly for that item. So the item is too difficult. High difficulty value index means, that item is an easy one, e.g., D.V = 0.80 it means 80% answered correctly for that item. So that item is too easy one. According to (Ebel, 1991) there are five standards for discriminating the value of items with respect to their evaluation as given in table 3.

Table 3. Standards for difficulty value.

Difficulty value	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

Discrimination power: To calculate the Discrimination Power the researchers used the following formulae:

Discrimination equation: D= $(p1\text{-}p2)/n1$

p1= the number of students who give right answer in high group, p2= the number of students who give right answer in low group, n1= the sample no. of high group or low group (Table 4).

Table 4. Item's discrimination power.

Item No.	Discrimination Value	Item No.	Discrimination value	Item No.	Discrimination value
1	.07	16	-.07	31	.33
2	.33	17	.47	32	.47
3	.13	18	.00	33	.33
4	.47	19	.13	34	.47
5	.47	20	.47	34	.53
6	.40	21	.47	36	.60
7	.20	22	.60	37	.47
8	.40	23	.60	38	.33
9	.40	24	.40	39	.53
10	.27	25	.00	40	.60
11	.53	26	.20	41	.53
12	.67	27	.13	42	.73
13	.47	28	.07	43	.47
14	.53	29	.53	44	.73
15	-.07	30	.07	45	-.07

General guidelines for discrimination: According to (Ebel, 1991) there are four standards for discriminating the value of items with respect to their evaluation as given in table 5.

Relationship between difficulty value and discrimination power: Both difficulty value and discrimination power are complementary not contradictory to each other. Both are considered to select good items. If an item has negatively discriminate or zero discrimination, was rejected whatever, the difficulty value is. On the basis of the above criteria items are acceptable in difficulty level as well as in discrimination power, except items no. 1, 3,15,16,18,25,27,28, 30 and 45. These items have been deleted because some of them are very difficult and have negative or negligible discriminating power.

Evaluation of test validity: A test is said to be valid if it measures what it has been to measure (Best, 1982). To determine the validity of the test, the researchers tested face validity, construct validity and discrimination validity.

Face validity or content validity: The content validity of the 'Reasoning Ability Test' was tested by more than 25 experts. It is evident from the assessment of experts that items of the test are directly related to the different dimensions of reasoning ability.

Construct validity: In order to find out the construct validity, the researchers calculated correlation between the score of each dimension and total score of the test (Table 6).

Table 5. Standards for discrimination value.

Discrimination Value	Item Evaluation
≥0.40	Very good item
0.30-0.39	Reasonably good but subject to improvement.
0.20-0.29	Marginal items need improvement.
<0.19	Poor items, rejected or revised

Table 6. Correlation between each dimension and total score.

Domain	One	Two	Three	Four	Five	Six
'r' values	0.751**	0.617**	0.525**	0.649**	0.739**	0.725**
Sig.	0.000	0.000	0.000	0.000	0.000	0.000

From the above table, it can be concluded that the correlation coefficient of all dimensions (.751, .617, .525, .649, .739, and .725respectively) is significant. This indicates that all dimensions are related to reasoning ability and the test has good construct validity.

Discrimination validity: To find out the discrimination validity of the items the researchers

used item analysis (difficulty level value and discrimination value). For knowing the level of discrimination validity for each dimension of the test, 't' test for two independent samples was used (high group and low group).

Finally the discrimination validity of whole test was

also determined by using 't' test. Discrimination validity for each domain and whole test is given in the table no. 7. It indicates that all 't' values are significant at level 0.01 and the means of high group are also higher than low group which support the high validity of reasoning ability test.

Table 7. Discrimination validity.

Dimensions	Group	N	Mean	Std. D	t	df	sig.
Analogical reasoning	High	30	7.07	1.23	5.90	58	.00
	Low	30	5.10	1.35			
Linear reasoning	High	30	3.67	1.24	4.68	58	.00
	Low	30	2.20	1.19			
Conditional reasoning	High	30	5.33	1.54	2.12	58	.00
	Low	30	4.53	1.22			
Deductive reasoning	High	30	7.13	1.17	6.12	58	.00
	Low	30	5.37	1.07			
Inductive reasoning	High	30	5.97	1.10	5.23	58	.00
	Low	30	4.00	1.74			
Cause & effect reasoning	High	30	3.93	1.62	4.41	58	.00
	Low	30	2.10	1.60			
TOTAL	High	30	33.10	3.99	9.27	58	.00
	Low	30	23.30	4.20			

Reliability of the test: The degree of consistency among test scores is called reliability. The reliability of the test was tested by calculating Alpha Cronbach Coefficient. The values of reliability coefficient for each domain test.

From the glance of table 8, all the values of reliability coefficient for each domain are highly significant. Thus reasoning ability test is a reliable test whose reliability is 0.71 and the reliability for each dimension is .65, .75, .63, .65, .73 and .71 respectively.

Table 8. Values of reliability coefficients for different dimensions.

Dimensions	Alpha value	Dimensions	Alpha value
Analogical reasoning	.65	Deductive reasoning	.65
Linear reasoning	.75	Inductive reasoning	.73
Conditional reasoning	.63	Cause & effect reasoning	.71
		Total Reliability	.71

The standards for interpretation of the test score: To categorise the students into different categories with respect to their reasoning ability the researchers used the standards calculated by using the Percentiles as given in table no. 9.

Final format of the test: Only 35 items related to six dimensions of RAT were selected in final format of the test. These include seven items for analogical reasoning, five items for linear reasoning, five items for conditional reasoning, five items for deductive reasoning, seven items for inductive reasoning and six items for cause and effect reasoning.

Table 9. Standards for categorization.

Category	Standard
Weak	0 –17
Acceptable	18–21
Good	22–23
Very good	24 – 28
Excellent	29– 35

RESULTS

After following these steps to construct the test and after analyzing the data from the first and the last application by using adequate statistical methods, it has been concluded that:

- The study has produced a reasoning ability test of secondary school students. This test includes (35) items which measures six types of reasoning ability i.e., analogical, linear, conditional, inductive, deductive and cause and effect reasoning.
- The test has been validated through content, construct and discrimination validity. The content validity has been evaluated by experts, construct validity has been calculated by Pearson's correlation. The correlation coefficients of all dimensions are (.751, .617, .525, .649, .739, and .725 respectively) which are significant. This indicates that all dimensions are related to reasoning ability and the test has good construct validity. The discrimination validity has been evaluated by 't' test for two independent samples (high group and low group). All 't' values are significant at level 0.01 and the means of high group are also higher than low group which support the high validity of RAT.
- The reliability of the test was tested by calculating Alpha Cronbach Coefficient. All the values of reliability coefficient for each dimension are highly significant. Thus reasoning ability test is a reliable test whose reliability is 0.71 and the reliability for each dimension of RAT is .65, .75, .63, .65, .73 and .71 respectively.
- To categorise the students into different categories with respect to their reasoning ability the researchers used the standards calculated by using the Percentiles. Students who get up to 17 points are considered weak, 18-21 are acceptable, 22-23 are good, 24-28 very good and 29-35 are excellent.

REFRENCES

Aggarwal, R. S. (2013). *A Modern Approach to Verbal Reasoning*. S. Chand & Company Ltd. Ramnagar New Delhi.

Alderson, J. C., & C. Clapham, D. W. (1995). *Language Test Construction and Evaluation.* Cambridge University Press.

Bandhana, & Sharma, D. (2012). A Study of Home Environment and Reasoning Ability among Secondary School Students. *Developing Country Studies,* 2 (1), 73-80. Retrieved from www.iiste.org.

Best, J. W., & Khan, J. V. (1982). *Research in Education*. Fourth Edition, Prentice Hall of India Pvt.

Cavallo, A. M. L. (1996). Meaningful learning, reasoning ability and students' understanding and problems solving topics in genetics. *Journal of Research in Science Education,* 33, 625-656.

Chen, T. C. (2000). Technology, reasoning, and freedom: Modern and postmodern conditions. *Taipei: Laureate* (In Chinese).

Ebel, R. L., & Frisbie, D. A. (1991*). Essentials of Educational Measurement.* (5th ed). Englewood Cliffs, NJ: Prentice Hall.

Fatima, Sk. & Rao, B. D. (2008). *Reasoning Ability of Adolescent Students*. Discovering Publishing House Pvt. Ltd. Darya Ganj New Delhi.

Jøsang, A. (2008). Conditional Reasoning with Subjective Logic. *Journal of Multiple-Valued Logic and Soft Computing,* 15(1), 5-38.

Jeotee, K. (2012). *Reasoning skills, problem solving ability, and academic ability: implications for programme and career choice in the context of higher education in Thailand*. Ph. D. Edu. Durham University.

Kemler, W. (1998). Sabbatical leave Report: *Suggested Guidelines for Selecting or Creating Program to Enhance Thinking and Reasoning Skills* (ERIC information No. ED33943.

Lawson, A. E., & Thompson, L. D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. *Journal of Research in Science Teaching,* 25, 733-746.

Leighton, J. P., & Sternberg, R. J. (2004). *The nature of reasoning*. New York, NY: Cambridge Uni. Press.

Shu, Y. M. (2000). Postmodern and Moral Education. *Bulletin of the National Institute of Educational Resources and Research Special Issue of Moral Education*.

Wu, M. (2001). The identification of reasoning skills mechanism, *Educational Research Monthly, 83, 72-93.*

Yenilmez A., Sungur, S., & Tekkaya, C. (2006). Students' achievement in relation to reasoning ability, prior knowledge and gender, *Research in Science & Technological Education,* 24:1, 129-138. Retrieved on 15/2/2013 from http://www.tandfonline.com.