Construction of an Achievement Test Measuring Academic Performance in Environmental Knowledge

Madhumala Sengupta[1], Mrinal Mukhopadhyay[2] and Pintu Kumar Maji[3]

[1]Department of Education, Calcutta University, Kolkata, India

[2,3]Nangi High School, Calcutta University, Kolkata, India

dr.msengupta@rediffmail.com, mrinal_bata@rediffmail.com, maji.pk@rediffmail.com

Environmental education is now a compulsory subject in school curriculum. The objectives are lofty. It is expected that along with acquiring of environmental knowledge the students will develop pro environmental attitude and problem solving skill. Testing of students environmental knowledge gleaned from school curriculum is one way of evaluating the whole programme of environmental education. However, teacher made tests are not often scientific and biases may inadvertently creep into them. So an achievement test on environment knowledge gathered from different school subjects namely physical science, biological science and geography was constructed following scientific procedures of test construction. The reliability of the test was found to be high (r = 0.89). The test was also successfully used in environmental research at the University of Calcutta.

Keywords: Achievement test, Environmental knowledge, Test construction

Introduction

The subject environmental education is taught compulsorily at the different stages of education. Although it is a separate disciplinary subject yet other school subjects specially science, contain components which are directly related to environmental issues. An achievement test constructed on the basis of such related content selected from science subject will help to assess the environmental knowledge of the students. Such a test, if constructed by adopting scientific procedure, will ascertain the students' grasp over environmental knowledge. Adequate knowledge base on environmental issues in turn is instrumental in developing environmental attitude, problem solving skill, affective values which are essential for environmental literacy. With this intention in mind the researchers have endeavored to develop a standardized scale.

Procedure of the development of achievement test of environment related knowledge

A number of scales on environmental attitude and awareness have been constructed for research purpose. The first scale in this respect was New Environment Paradigm Scale (Dunlap & Van Liere, 1976). It was revised by other authors. In our country many such tests measuring environmental awareness was constructed (Sengupta, 2005, EAAM by Jha 1998, 2004). Morgil, Arda, Secken, Yavuz and Ozyalcinsky (2004) studied chemistry education and environmental knowledge and constructed a test for this purpose.

However, in this case a new approach has been adopted in which the test items were selected from school subjects on the basis of their environment related loading. Of course, there are standardised tools to measure the levels achievement in physics, chemistry and other subjects. But in this study a new approach in test construction has been adopted in which case the test will measure only those aspects of knowledge that are related to environmental issues. A set of standards for educational tests has been devised by a permanent commission COTAN in the Dutch Institute of Psychology (Evers et al., 2002). These are:

- Purpose and scope, Quality of test material, Quality of the manual, Norms, Reliability.
- Content and Construct validity, Criterion validity.

Test purpose

A number of taxonomies of purposes of educational tests are available (Linn & Gronlund, 2005; Mehrens & Lehmann, 1991). This particular test has been constructed with sole purpose of assessing the environmental knowledge hence the curricular domain is to include knowledge, understanding and comprehension. The aim of any test is to infer whether the domain of ability or proficiency has been achieved by the test takers. Millman and Green (1989) suggested that the test content should have the following five features:

- *Sources of the test content*- It has already been referred that the test content will be drawn from curricular domain having cognitive implication without any emphasis on competency domain or the domain of future criteria.

— *Dimensionality of test*- Dimensionality is about homogeneity or heterogeneity of a test. In this case the test has homogeneous dimension as it measures environment related knowledge though selected from different subjects.

— *Domain versus norm referenced test*- The test that has been constructed would measure absolute performance of each individual student and hence domain specific.

— *Bandwidth versus fidelity*- The test is narrow in scope lacking breath of content coverage but hopefully be more reliable.

— *Content distribution*- The items have been selected across the content domain having environmental implications.

A panel of three experts were requested to identify the major areas along which environmental concern, content and knowledge could be assessed. These three experts are teaching in secondary teachers' training college and University of Calcutta respectively life sciences, physical sciences and geography as a method along with environmental education. They were requested to suggest the major areas of the respective subjects namely—-*life sciences, physical sciences* and *geography* which have maximum concern with environment and growing environmental problems. They were particularly requested to keep in mind that the examinee would be students who had already completed the syllabus of class-VIII. It was decided that the aim of the test would be to assess the respondents' general and specific knowledge about environment, understanding of the contents and issues and their ability to choose appropriate skill from given option to solve the specific environmental problems. They were also duly reported that the test would be multiple choice types and there would be only single right answer along with three distracters. For the assessment of students' knowledge, understanding and skill in a given situation, the experts suggested the following areas or dimensions:

Life sciences—-Structure of ecosystem, Component of ecosystem, Interaction between different components, habitat study, vectors and pest, pollution and bio-hazards bad habits like smoking etc and their effects on human health.

Physical science—-Pollution and pollutant, lifestyle, consumption behavior and environmental degradation. Industrialization and environment, solid waste management, engineering apparatus/gadgets use to check pollution.

Geography—-Climate study, Resource study, study of natural disaster, global warming, ozone layer depletion and acid rain, urban planning.

After the content areas have been identified, the next step was preparation of a *table of specification*. The table shows the relation between the content and the different dimensions of cognitive bahaviour. The taxonomy of different levels of cognitive behaviour is based on well known taxonomy of Bloom (1976). A shortened version of Bloom's original six hierarchical cognitive domains is applied here i.e. Knowledge, understanding (information competence) and application (Inferential reasoning). The following Table 1 of specification served as blue print of test construction and was used in item writing scheme.

Content Areas	Factual Knowledge	Information Competence	Inferential Reasoning	Total no. Of Items
Life Science	10	10	12	32
Physical Science	8	8	6	22
Geography	9	5	6	20
Total	27	23	24	74

Table 1: Specification for item writing

Constructing test items

This is the most important part of test construction. Constructing test items is both an art and a science and requires creativity on the part of the test developers. The important issues in this respect are:

— Range of item difficulty

— How many items should be included in the test

— How many initial items should be chosen

— What types of cognitive domains should be measured

— Since it is an environmental knowledge test whether affective and psychomotor domains should also be measured

— Lastly what kind of test format should be used?

Range of item difficulty should be sufficiently varied in order to differentiate between high achievers and low achievers. Initially in this case, excess number of items was selected in order to ascertain that only those items are retained which have optimum difficulty range. The length of the test is to be considered. If it is too long then the students will feel fatigued and too short a test will not have sufficient reliability. It was decided that if approximately fifty items from physical science, biological science and geography are included in the test then it will be of optimum length.

Item format and item writing

Selection of item format should be based on quality criteria, taking into account the issues of validity, reliability, appropriateness, feasibility, transparency etc. Haladyna (1992)

mentioned two types item format namely selected –response format and constructed response format. In this case selected response format is used. In selected response format the students have to choose an alternative from some pre-specified alternatives. This format comprises different versions like true-false, multiple choice, and matching items. Among the different versions of selected response version, multiple choice type items were constructed for this test. Following precautions were taken before the items were written

- Items were accurate and valid
- There was only one correct answer
- Alternatives were mutually exclusive
- Wordings of the stem and response alternatives were unambiguous
- All options were plausible and attractive to the students
- No clue was given for correct alternatives (Scheerens, Glas, & Thomas, 2007)
- "None of the above" and "All of the above" options were not used.
- Use of non-functional word was avoided
- Stems were not taken verbatim from the textbooks
- Distracters are chosen with utmost care

Multiple choice items have some short comings. For example it is not suitable for testing proficiency requiring writing skill, mathematical reasoning or real life performance. Yet this format was selected because wide content coverage is possible through it. In the original try-out pool seventy four (74) items were included because it was expected that many items will have to be discarded or revised as the development of the test proceeds. The next step in test construction is to use item analysis.

Item Analysis

Item analysis is a family of statistical procedures to identify the best items. A thorough item analysis can be conducted by using the various methods like; Item-difficulty index and Index of item discrimination. Item difficulty is the proportion of examinees who correctly answer the item when the test is applied as a trial. It is necessary to find out the difficulty index items so that the "too easy" or "too difficult" items are discarded in the final test. Generally items having difficulty index values between 0.3 - 0.7 are acceptable in the final test. However, the selection criteria may vary depending on the situation in which the test is used. As the purpose of the test was to conduct further research on environmental education, the above mentioned range is acceptable.

An item discrimination index is a statistical technique of finding out how well an item is able to differentiate between high scorers and low scorers. In a normally distributed score the upper 27% and lower 27% of scores are taken into account to calculate discrimination index. The 'd' or discrimination index of an item is $d = U\text{-}L/N$ where U is the number of students who correctly answered the item in the upper range and L is the number of students who correctly answered the item in the lower range and N is the number of examinees in upper or lower group. A positive value of d is required and preferably it should be closer to 1.

Reliability and validity of the test

Reliability implies the consistency in measurement. It indicates the stability of the test scores. There are different methods of assessing reliability of a test. In this case coefficient alpha formula (KR-20) and Spearman-Brown prophecy formula were adopted to determine the reliability of the whole test. On the basis of Item Difficulty and Item Discriminatory Indices as mentioned above, reliability of test by KR formula (KR-20) was found to be 0.88.

The reliability also calculated by Split Half Method according to Spearman-Brown prophecy formula. The procedure is to make up two sets of scores by combining alternate items in the test. The first set of scores represents performance on the odd numbered items—5, 7, 13 etc. and the second set of scores, performance on the even number of items—2, 4, 6, 8, etc. The odds-evens split is the one most commonly used method. It is noteworthy to mention here that the split-half method is regarded as best of the methods for measuring test reliability and the main advantage is that all data for computing reliability are obtained upon one occasion; so that the variations brought about differences between the two testing situations are eliminated (Garret, 2007). Reliability by split-half method was found to be 0.89. So it is evident that the Test is reliable.

Validity is the evaluative judgement of the degree to which the tests measures what it purports to measure. Good item validity is in itself a guarantee of test validity and therefore. The questionnaires were/test was considered valid. Since all the items were scanned and rated by the expert, the content validity was ensured. The criterion related validity of the test was determined by comparing the scores of the respondents of environment knowledge test and their scores of annual exam in three school subjects namely life sciences, physical sciences and geography. The criterion related validity signifies the performance of the respondents on some out come measure. In this case the out come measure is the score of annual examination in three above mentioned school subjects.

Instructions

The test of environmental knowledge was provided with specific instruction. In the front page of the booklet following general instructions were printed.

"In this test paper you are going to get seventy four (74) questions from the syllabus of class VI to class VIII covering the different aspects of environment. Answer the questions without any hesitation. Each of the questions has same weight. Choose the right answer by making a circle (O) among the given alternative answer. If any early response found to be wrong make it dark (O) and mark the next choice. Try to answer all the items".

Time allotted for the test was 45 minutes. On the basis of above mentioned techniques, the test was given final shape by retaining only those items having optimum difficulty and discriminatory indices. The selected items were further analyzed to find out detailed responses of the subjects. It implies the patterns of responses regarding the distracters apart from the correct answers. It was found that no distracter has 0 value, and the distribution of frequency of responses in respect of distracters was more or less equitable. The test (see Appendix) was ultimately used successfully for pursuing research work on the methodology of environmental education.

References

Bloom, B. (1976). *Human characteristics and school learning*. New York: McgrawHill.

Dunlap, R.E., & Van Liere, K.D. (1976). The new environmental paradigm. A proposed measuring instrument and preliminary results. *Journal of Environmental Education*, 9, 10-19.

Evers, A., Vilet, Mulder, J.C. van, Resing, W.C.M., Starren, J.C.M.G., Alphen de Veer, R.J., van, & Boxtel, H. van (2002). *COTAN: Testboek voor het onderwijs*. Amsterdam: NDC Boom.pag.

Garrett, H.E . (2007). *Statistics in psychology and education*. New Delhi: Paragon International Publishers.

Haladyna, T.M. (1992). Context dependent item sets. *Educational Measurement: Issues and Practice*, 11, 21-25.

Jha, P.K. (1998). *Environment awareness ability measure*. National Psychological Corporation Agra.

Jha, P.K. (2004). *Environment awareness ability measure*. National Psychological Corporation Agra.

Linn, L.R., & Grosnlund, N.E., (2005). *Measurement and assessment in teaching.* Delhi: Pearson Education Pvt. Ltd.

Mehrens, W.A., & Lehmann, I.J. (1991). *Measurement and evaluation in education and psychology.* New York: Holt, Rinehart and Winston.

Millman, J., & Green, J. (1989). The specification and development of tests of achievement ability. In R.L. Linn (Ed.), *Educational Measurement* (3rd edn.), 335-366. New York, NJ: American Council on Education and Macmillan.

Morgil, I., Arda, S., Secken, N., Yavuz, S., & Ozyalcinskay, O. (2004). The influence of computer assisted education on environmental knowledge and environmental awareness. *Chemistry Education: Research and Practice*. 5(2), 99-110.

Scheerens, J., Glas, C., & Thomas, S. (2007). *Educational evaluation assessment and monitoring*. New York: Taylor and Francis.

Sengupta, M. (2005). *Environmental awareness of environmentally active and passive students in relation to their motivation and academic achievement*. Unpublished Ph.D thesis. Department of Education, Calcutta University.

Appendix

Table shows the item chosen (high lighted by * sign) for final test on the basis of difficulty and discriminatory indices (N = 230)

LIFE SCIENCES			PHYSICAL SCIENCES			GEOGRAPHY		
ITEM NO.	ITEM DIFCLT	DISCR. INDEX	ITEM NO.	ITEM DIFCLT	DISCR. INDEX	ITEM NO.	ITEM DIFCLT	DISCR. INDEX
1*	0.5	0.59	1*	0.4	0.44	1*	0.4	0.59
2*	0.5	0.5	2*	0.3	0.45	2*	0.5	0.42
3	0.4	0.01	3*	0.6	0.45	3*	0.6	0.73
4	0.9	0.16	4	0.3	0.27	4	0.8	0.69
5	0.4	0.27	5	0.2	0.16	5	0.8	0.4
6*	0.5	0.44	6*	0.3	0.55	6	0.3	0.14
7*	0.3	0.44	7*	0.7	0.5	7*	0.7	0.4
8*	0.7	0.44	8*	0.5	0.53	8*	0.3	0.53
9*	0.4	0.48	9	0.4	0.31	9*	0.7	0.45

10*	0.6	0.36	10*	0.4	0.47	10	0.2	0.03
11	0.2	0.17	11	0.7	0.23	11*	0.4	0.56
12*	0.5	0.55	12*	0.6	0.48	12*	0.4	0.37
13	0.4	0.21	13*	0.6	0.57	13*	0.6	0.55
14	0.8	0.15	14*	0.5	0.42	14*	0.6	0.47
15*	0.6	0.45	15*	0.4	0.39	15*	0.7	0.53
16*	0.5	0.74	16	0.2	0.21	16*	0.5	0.79
17	0.8	0.42	17*	0.6	0.39	17*	0.6	0.36
18*	0.3	0.5	18*	0.6	0.59	18	0.8	0.36
19*	0.7	0.4	19*	0.5	0.59	19*	0.5	0.59
20*	0.3	0.45	20*	0.3	0.34	20*	0.6	0.53
21*	0.4	0.3	21*	0.7	0.52			
22*	0.4	0.47	22	0.3	0.32			
23*	0.7	0.48						
24	0.9	0.24						
25*	0.5	0.58						
26	0.2	0.13						
27	0.5	0.27						
28	0.8	0.37						
29*	0.7	0.5						
30	0.9	0.27						
31	0.8	0.36						
32*	0.7	0.52						