Development of Achievement Test on Statistical Application and Interpretation for Assessing Learning of Educational Researchers Through Module

S. K. TYAGI*

Abstract

In the backdrop of University Grants Commission's (UGC) decision to relax the eligibility criteria for research guides, need was felt to augment the educational researchers' skills in statistical application and interpretation through the module. Apart from developing a module on the subject, it was planned to develop an achievement test to assess the effect of treatment in terms of researchers' understanding of the statistical application and writing interpretation of the results. Accordingly, an achievement test was constructed in two parts. Part A focused on the assessment of understanding of statistical tests and appropriate research situations, it comprised of 50 multiple choice type test items. Part B, consisting of 10 free response test items tested researchers' skill of interpreting results of the statistical tests. The tests were reviewed by the experts of research methodology and testing. The reviewed tests were tried out on a small sample of 18 researchers of different universities participating in two research methodology workshops at University of Allahabad and Regional Institute of Education (RIE), Ajmer respectively. Response analysis of the test items was performed. The feedback obtained from the response analysis led to the needed item modification and item replacement. Cronbach's a was computed following the final administration of the test.

**ICSSR Senior Fellow*, School of Education, Devi Ahilya Vishwavidyalaya, Indore - 452001

Introduction

Need of Improving Research Skills of Practitioners through the Module

UGC, through its notification, 'University Grants Commission (Minimum Standards and Procedure for Award of M.Phil./Ph.D. Degree) Regulations, 2016', decided to relax the criteria for recognising a research guide. This led to a steep rise in the number of recognised guides and by implication, the number of Ph.D. candidates working/aspiring to work under their supervision would also increase. The scenario of quality of educational research from the point of view of statistical application has not been satisfactory, as evidenced by Govil et al. (2015), Tyagi (2019) . In the wake of these developments, it is feared that unless the research knowledge and skill repertory of the research scholars and the research guides are adequately strengthened, the quality of the research work carried out in the universities may go down further.

However, there are ways to handle this situation. One way is to offer refresher courses in research methodology and organise workshops on statistical application and interpretation. Another alternative is to develop open online courses, Massive Open Online Courses (MOOCs) on statistical application and interpretation, such as launched on UGC digital platforms like Swayam or the e-PG pathshala under National Mission on Education using ICT (NME-ICT). The problem with video lectures, however, is that most of the developed video lectures are in English medium with an accent unfamiliar for most of the viewers. Besides, video lectures are basically passive strategies that lack in active learning tasks by the learners. The required tasks for learner engagement for skill development through practical work is missing in such strategies. Thus, the opportunities for the development of skills of statistical application and interpretation through video lectures are somewhat limited. Use of learning modules in Hindi may offset these problems to some extent for a large section of the researchers' population.

A Modular Approach to Learning

A module is a self-learning material defined differently by different authors. Houston (1972) defined a module as a 'set of experiences designed to facilitate the learners' demonstration of specified objectives'. According to Khasnavis (1983), a module is defined as an instrument which 'systematises the learning process of students and holds the student accountable for his learning experiences'.

Thus, a module includes a set of activities intended to facilitate the learners' achievement of specific objectives. It is relatively self-contained and self-pacing unit of instruction designed for

specific purposes and as a part of a comprehensive instructional system.

Rationale for Experimenting with Modular Learning

The modular approach to learning has been studied in the context of a few subjects at school level as also at the teacher education level.

A study by Kohal (1999) revealed that the students of Class XI taught through mastery teaching strategies attained geography concepts better than the students taught through non-mastery strategies. Londhey (2007) studied the effectiveness of the modular approach for teaching science to Class IX students in terms of their achievement. The finding of the study was that the modular group was significantly effective in terms of achievement of students more than the other group.

Mollykutty (1991) studied the effectiveness of modular approach in teacher education and found that achievement through a modular approach was significantly higher than the existing approaches adopted in teacher education institutions. Joshi (1999) found that the module was effective in terms of the achievement of B.Ed. students of Devi Ahilya Vishwavidyalaya in Educational Technology. Maharana (2011) conducted a study which found that the module with and without jerk technology was effective in terms of achievement of students in environmental education when groups were matched with respect to pre-achievement in environmental education. Shinde (2007) found video instructional material to be effective in terms of achievement of Post Graduate students in Research Methodology and Statistics, while Sultan and Tyagi (2018) found modular approach in Research and Statistics to be significantly more effective than e-lecture approach.

Furthermore, modules have been used to enhance the communication and management skills of the heads of educational institutions. Chopra (2002) undertook a study on self-instructional module for enhancing the communication skills of college principals. It was observed that the communication module has definitely been useful for college principals in enhancing their communication skills and competence to optimally perform at work. Shetty (2004) conducted a study on self-instructional module on staff development for secondary school principals. The study concluded that— (i) Modules helped the principals to enhance their abilities of organisational management, (ii) Principals were able to perform better in encouraging the staff, increasing staff participation, effective time management, and working effectively under pressure by studying through the module.

Learning modules, particularly in the Hindi language, have not so far been used for developing the achievement of educational

researchers in the field of statistical applications and interpretation.

Assessment of Achievement of Modular Learning

The present study focussed at experimenting with teaching statistical application and interpretation to researchers in education through the module. Apart from conducting the experiment, the tasks involved in the study were based on developing— (i) a module consisting of 8–10 capsules on various topics under the focal areas of statistical application and interpretation, and (ii) an achievement test on statistical application and interpretation. The present paper is confined to the second objective listed above.

Development of Achievement Test on Statistical Application: Test A

The study intended to develop module on statistical application and interpretation for the use of research practitioners in education and study their effectiveness in terms of conceptual understanding of the tests applied and the skill of writing interpretation of results. Accordingly, a criterion-referenced test was proposed to be used both at the pre- and post-experiment stage in two parts, namely: Part A, covering statistical application, and Part B on writing interpretation of the results. The first part of the test was supposed to be an objective type test focussed on the understanding of the concepts underlying the application of statistical tests. The second part of the test was to be based on the interpretation of the results of test of hypotheses. Part B of the test was designed in the subjective format allowing the subjects to write their responses freely. The process of the development of each of these parts of the tests is given below in detail .

Preparation of the initial draft

An objective type test comprising of 50 multiple response type questions to be answered in about an hour, was conceptualised for this purpose. First of all, an initial draft was prepared covering the following content areas:

Table 1
Distribution of the Content Areas in the Test Along with Weights

Content area	Serial no. of Items	Total no. of Items	Weight in Percentage
Assumptions underlying tests and their testing procedure	1, 2, 3, 4, 5, 7, 8, 17, 18, 19, 21, 22, 25, 42, 44, 46	16	32
Selection of appropriate tests of hypotheses	9, 15, 23, 24, 26, 27, 28, 30, 33, 36, 39, 43, 45	13	26

Interpretation of results	6, 13, 14, 16, 20, 31, 32, 34, 35, 37, 38, 40, 41	13	26
Post significance analysis	10, 11, 12, 29	04	08
Effect size measures	47, 48, 49, 50	04	08
Total		50	100

The initial draft of the test was mailed to two subject experts—one from the area of research methodology and the other from testing area, for their expert views on the test. The draft was reviewed in light of the received comments.

Small sample try out

The preliminary draft was subjected to empirical testing for getting feedback on the quality of the questions framed. The sample for small group try out comprised of 18 UGC JRF and Research scholars selected purposively from two institutions as per details given below.

Table 2
Sample for Small Group Tryout (N=18)

University/ Institution	Subjects	No. of Subjects
University of Allahabad	UGC research fellows	10
Regional Institute of Education, Ajmer	Research scholars	08

Analysis of the results of small sample tryout

The test performance of the subjects was analysed using SPSS. Various test parameters were observed, which are as under:

Mean = 10.89 Median = 11.00
Variance = 36.69 SD = 6.06
Minimum = 1 Maximum =21
Skewness(Sk) = 0.158
Standard error = 0.536
Kurtosis (Ku) = -0.84
Standard error = 1.3

Table 3
Shapiro-Wilk Test of Normality

S-W Statistics	df	Sig.
0.958	18	0.535

The results of the Shapiro-Wilk normality test (SW = 0.958, p= 0.535 >0.05) show that the distribution of scores on the test was quite normal. The descriptive statistics related to the test show that the value of skewness is less than its standard error, and hence can statistically be regarded as 0. Likewise, kurtosis too does not deviate substantially from the corresponding value of a normal distribution, being within ± 1.96 of its standard error. The obtained scores ranged from a minimum of 1 to a maximum of 21.

The test items were further subjected to response analysis to detect the cases of probable guessing,

leading to the identification of items needing item reformulation, modification or replacement.

(i) Test items with no response or no correct response

The item responses were tabulated and analysed. Out of 50 questions, there were 4 items which elicited no correct response at all. The serial numbers and the content areas for these items are presented below:

Table 4
Items without a Single Correct Response

Item Serial	Contents Covered
7	Data Transformation
11	Post-hoc Test in Case of Heterogeneity
26	Statistical Test Suitable for 2-Groups Pre-test Post-test Design
40	Interpretation of Table of Cell Means in Factorial ANOVA

All these items were centred on the concepts which are normally not included in the prevailing PG and post PG course curriculum of research and statistics. Similar, was the situation for those items which registered just one correct response each from the entire sample. The details are presented below.

(ii) Test items with one correct response each

Three items given below were answered correctly by just a single respondent. The content areas tested by the item are mentioned against each item.

Table 5
Items with just a Single Correct Response

Item Serial	Contents Covered
21	Assumption of Homogeneity of Regression slopes
43	Non-parametric Paired Samples Test
49	Effect Size for Two-way ANOVA

The responses to these items were clearly not influenced by pure guessing, implying that the items possessed good quality distracters. These items were, therefore, retained in the final draft without any change.

(iii) Test items with an unexpectedly high correct response probably due to guessing

Upon analysis of the frequencies of the correct response and the contents covered, it was revealed that the following items elicited unexpectedly high correct response as the concepts on which the items were based were almost new to the respondents. Below are given the relevant information regarding these items.

Table 6
Items with an Unexpectedly High Correct Response Rate

Item Serial	Contents Covered	Correct Response	%
8	Levene's Statistis-Conceptual Base	4*	22
10	Most Liberal Post-hoc Test	5	28

47	Effect Size for t-test	4	22

**Maximum* correct responses for any item of the entire test were 9/18.

Item modification

The items which received relatively high correct response needed further investigation. The occurrence could be due to inadvertently suggestive nature of their stems making it easier to guess the key. Alternatively, it might be caused by the inferior quality of the distracters. Thus, after taking all possibilities into account, the following modifications were made in the test items. Table 7 shows item serial, the contents of the original item, item-wise suggested modifications and the probable reason thereof.

Table 7
Items with the Modifications or Replacements Applied along with the Reason

Item Serial	Original	Modification Proposed	Reason
8	Levene's Statistics	Levene's Statistics for Homogeneity of Variance	Will add to the value of the distracter containing the word 'Variance'
8	Group Means	Group Medians	Inadvertent typological error making the two distracters indistinguishable
10	Most Liberal	Most Conservative	The word 'least' was perhaps suggestive of the 'liberal' test
47	Effect Size Measure for t-test....Cohen's d	The item was replaced with no reference to 'd'	Since the formula for t-test contains statistics 'd', the key [Cohen's d] was suggestive
50	Cohen's d of 0.8 means....large effect size in case of t-test	The item was replaced with a new question	Respondents perhaps thought the value of 0.8 (as in correlation coefficient 'r') is large

All the remaining items needed no modification as the respondents were expected to learn something regarding these areas in their PG, M. Phil. or Pre-Ph.D. coursework, and hence were retained as such in the final version of the test.

Final Administration of the Test

The final version of Test-A was administered on 33 participants of a National Workshop on Statistical Application and Interpretation held at Bareilly. The participants included research fellows and

research scholars from M. J. P. Rohilkhand University, Bareilly; Kumaun University, Almora campus; Dr. Harisingh Gour University, Sagar and C. C. S. University, Meerut. Besides researchers, the participants also included the faculty members of the department of education of Jyoti College of Management and Sciences, Bareilly affiliated to Mahatma Jyotiba Phule Rohilkhand University, Bareilly.

Analysis of the Results of the Final Administration

The test performance of the subjects was analysed using SPSS. Various test parameters were observed, which are as under:

Mean = 10.45 Median = 11.00
Variance = 18.13 SD = 4.26
Minimum = 1 Maximum = 18
Skewness = -0.163 Standard error = 0.409
Kurtosis = -0.549 Standard error = 0.798

Table 8
Shapiro-Wilk Test of Normality

S-W Statistics	df	Sig.
0.981	33	0.822

The results of Shapiro-Wilk normality test show that the distribution of scores on the test was quite normal (SW = 0.981, p= 0.882>0.05). The descriptive statistics related to the test shows the skewness to be around 0 and kurtosis too not deviating substantially from the corresponding value of a normal distribution. The obtained scores ranged, now from a minimum of 1 to a maximum of 18. The variance of the scores in the final administration reduced considerably (36.69 to 18.13) indicating a reduction in scores by guessing and improvement in the quality of the test items.

Reliability analysis

The scores of the final administration of the test consisting of 50 items were subjected to reliability analysis, the results of which are presented below in Table 9 and 10.

Table 9
Case Processing Summary

Cases	N	%
Valid	33	100.0
Excludeda	0	.0
Total	33	100.0

Table 10
Reliability Statistics

Cronbach's Alpha	N of Items
0.624	50

The results reveal that the reliability coefficient of the test administered on a sample of 33 respondents was 0.624, which is quite close to the acceptable value of 0.7.

Development of Test of Statistical Interpretation: Test B

As indicated earlier, another part of the test (Test B) was planned to assess the skill of research practitioners in writing an interpretation of the output of statistical tests. It was decided to

have it as a free, extended response (commonly known as essay) type test consisting of 10 items to be answered in about one hour. As with Test A, an initial draft was constructed, to begin with. Questions cutting across all the statistical tests to be taught at the M.Phil. or Ph.D. coursework level were included in the test. Enough space for writing the response was provided on the test paper itself. The areas covered in the preliminary draft of the test are enumerated below.

Table 11
Item-wise Content Areas in the Preliminary Draft of Test B

Test item	Area Covered
1	Directional Hypothesis
2	Significant Interaction
3	Tukey's Test
4	Normality of Test Scores
5	Homogeneity of Variance
6	Homogeneity of Regression Slopes
7	Kruskal Wallis test
8	2-Samples Chi-square
9	Regression Analysis
10	Wilcoxon Test

SMALL SAMPLE TRY OUT

The preliminary draft of this test was also subjected to empirical testing for getting feedback on the quality of the questions framed. The sample for the small group tryout comprised of the same 18 UGC JRF or Research Scholars who were also administered in the first part of the test, i.e., Test A (refer to Table 2 for details). The responses were analysed with a view to improve the items, the results are discussed below.

Analysis of the Test Performance

An analysis of the test responses led to the following observations:

- Most of the researchers could only attempt questions on t-test, ANOVA, regression, normality, and homogeneity of variance. Regarding these, they could interpret the significance of the test statistics in the questions through p-values. They, however, were unable to write proper interpretation because of the lack of skill of formulating hypotheses commensurate with the statistical tests.
- Many of them just repeated the information from the tabulated output in the text form without further knowing how to answer the research question posed in the given situation.
- None of the research scholars attempted questions on interpretation of the results of non-parametric tests.
- The output of Tukey's test in the form of homogeneous subsets proved to be too tricky for the respondents.
- The output regarding chi-square test received no response at all. Similarly, the respondents did not have any idea regarding the residual analysis performed in view of a significant chi-square value of the test.

Modification in the test items

Insights into the responses led to a modification in the questions. Also, some of the items got dropped and replaced by the new items. The following table gives an idea of the modifications along with reasons.

Table 12
Changes made in the Revised Draft of Test B with Reasons thereof

Test item Serial	Test Item Serial Revised	Area Covered	Decision	Reason for Change
1	2	Directional Hypothesis	Retained	
2	3	Interpretation of Significant Interaction Through Graph	Retained	
3	4	Tukey's Test: Homogenous Subsets	Retained	
4	5	Normality of Test Scores	Retained	
5	1	Homogeneity of Variance Test	Replaced (by item on Writing Objective/ Hypothesis of Two-way ANOVA)	To avoid duplicity of items on assumption testing
6	6	Homogeneity of Regression Slopes	Replaced (by an item on One-way ANCOVA With one Covariate)	To avoid duplicity of items on assumption testing. Moreover, ANCOVA was not represented in the test
7	7	Kruskal Wallis Test	Retained	
8	8	2-Samples Chi-Square Test	Modified to include the Interpretation of Chi-square Value Along With Residual Analysis	Added the task of Stating and Testing of the Hypothesis (of Independence of Attributes)

9	9	Regression Analysis	Retained	
10	10	Wilcoxon Test	Retained	

Development of the Response Sheet

Since Test A was reusable, responses to it were supposed to be recorded on a separate sheet called, response sheet. The response sheet comprised of general information regarding the respondents such as, name, university or institution and the class, designation (in case of a faculty), apart from instructions to mark their responses on the response sheet. After tryout of Test A, it was decided to add one more entry under the title 'Subject,' so as to permit the selection of subjects from the field of education by a mixed group of researchers of social sciences.

Likewise, the entries printed on Test A were also added to the cover page of Test B. This was done to facilitate the identification of the respondents answering Test B from the response sheet of Test B itself.

Concluding Remarks

Thus, the present work related to the development of an Achievement Test on statistical application and interpretation for research practitioners in education offered a unique challenge on a couple of grounds. Firstly, the test was based on contents, some of which were new to the respondents unless taught to them. This led to a situation of almost zero variance in obtained scores on such items, thus not permitting item analysis using item discrimination index and item difficulty value. This, also, partly explains a nearly acceptable value of reliability coefficient Cronbach's α = 0.624.

Secondly, a part of the test was a free response (as against fixed response) or essay type, designed to test the interpretational skills of the researchers. The second part of the test, too, was not amenable to traditional item analysis and hence, only descriptive response analysis could be performed on it.

Acknowledgements

The author thankfully acknowledges ICSSR, New Delhi, for supporting this project under Senior Fellowship scheme.

References

Chopra, N. 2002. Development of Self Instructional Module to Enhance Communication Skills of College Principals. Unpublished doctoral thesis. M. S. University, Baroda.

Govil, P., N. A. M. Quasem. and S. Gupta. 2015. Evaluation of Statistical Methods used in Ph.D. Theses of Social Science in Indian Universities. *International Journal of Recent Scientific Research*. Vol. 6, No. 3. pp. 2926–2931.

HOUSTON, W. R. 1972. Developing Instructional Modules. Houston University, Houston.

JOSHI, KAMAKSHI. 1999. Development of Module on Educational Technology and its Effectiveness and Comparison with Traditional Method in terms of Scholastic Achievement at B. Ed. Level. Unpublished doctoral thesis. DAVV, Indore.

KOHAL, V. 1999. Effectiveness of Self Learning Modules on Achievement in Geography in Relation to Mastery and Non-Mastery Teaching Strategies, Intelligence and Study Habits. Unpublished doctoral thesis. Panjab University, Chandigarh.

KHASNAVIS, P. K. 1983. *Teaching of Social Studies in India.* Abhinav Publications, New Delhi.

LONDHEY, U. 2007. A Study of Effectiveness of Modular Approach for Teaching Science to Class IX Students in terms of their Achievement and Reactions towards Modular Approach. Unpublished doctoral thesis. D.A.V.V., Indore.

MAHARANA, N. 2011. Comparative Effectiveness of with and without Jerk Technology Module on Environmental Education in terms of Achievement in Environmental Education on B.Ed Trainees. Unpublished doctoral thesis. D.A.V.V., Indore.

MOLLYKUTTY, T. M. 1991. Effectiveness of Modular Approach for Teaching Education and Requisites for Implementation. Unpublished doctoral thesis. University of Kerala.

SHETTY, A. D. 2004. Development of Self Instructional Module on Staff Development for the Secondary School Principals. Unpublished doctoral thesis. M. S. University, Baroda.

SHINDE, L. 2007. Effectiveness of Video Instructional Material on Research Methodology and Statistics in terms of Achievement in Research Methodology and Statistics. Unpublished doctoral thesis. DAVV, Indore.

SULTAN, MUDASIR AND S. K. TYAGI. 2018. Comparative Effectiveness of Modular and E-Lecture approaches for Learning Educational Research Concepts by P.G. and Post P. G. Students. *Scholarly Research Journal for Humanity Science and English Language*. December–January. pp. 6976–6989.

TYAGI, S. K. 2019 [IN PRESS]. Statistical Audit of Data Analysis of Educational Researches. *Indian Educational Review*. Vol. 57, Issue 1.

UNIVERSITY GRANTS COMMISSION. 2016. UGC (Minimum Standards and Procedure for the Awards of M. Phil./Ph.D. Degree) Regulations. UGC, New Delhi.