DOCUMENT RESUME

ED 242 684 SP 024 076

AUTHOR Burke, J. Bruce; VanSusteren, Timothy J.
TITLE Modeling Teacher Education Research in an Introductory Teacher Education Course: A Case Study.
PUB DATE 30 Jan 84
NOTE 21p.; Paper presented at the Annual Meeting of the Association of Teacher Educators (64th, New Orleans, LA, January 28-February 1, 1984).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classroom Research; Educational Psychology; *Educational Research; *Education Courses; Education Majors; *Experiential Learning; Higher Education; Multiple Choice Tests; *Objective Tests; Preservice Teacher Education; *Questioning Techniques; Research Methodology; Student Attitudes; Test Construction; Test Reliability
IDENTIFIERS *Alternate Choice Questions

ABSTRACT

A study, conducted in a basic educational psychology course which serves as a gateway course to a teacher education program at Michigan State University, focused on the development of new knowledge while it taught about the educational research process and the subject being studied. A new format for objective testing, alternate-choice questions, was studied. Specific study project goals were to: (1) study the usefulness of a novel test-item format to evaluate student achievement; (2) increase the reliability and validity of the course's unit examinations and final examinations; (3) collect information useful in evaluating course effectiveness, student learning, and teacher performance; (4) facilitate student learning of practical and sound measurement theory and techniques through investigation of an important measurement question; and (5) model the spirit of inquiry, impressing on preservice teachers the importance and value of experimentation with new techniques and the necessity of continuous evaluation of both new and well-accepted practices. Study results showed that Alternate Choice (AC) items were slightly less difficult, discriminating, and reliable than Multiple Choice (MC) items; that students were able to respond to twice as many AC as MC items in a given time; and that students viewed the course testing positively. (JMK)

MODELING TEACHER EDUCATION RESEARCH IN AN INTRODUCTORY TEACHER EDUCATION COURSE: A CASE STUDY

Presented at: Association of Teacher Educators Convention

January 30, 1984

New Orleans

For Private Distribution Only

J. Bruce Burke

Michigan State University

and

Timothy J. VanSusteren

City of Milwaukee

The hallmark of a professional, no matter what the field, is the study and improvement of one's professional practice. Teacher educators have only recently begun to study systematically their own behaviors, practices, and techniques. As a profession we have been engaged for some time in the study of teachers and teaching. The results of such research has positively affected the content and style of our own teaching. It is common now to find teacher education courses which emphasize the importance of teachers becoming students of their own practices. At M.S.U. for example, we have "teacher collaborators" as full members of research teams. It is exciting to find teachers become involved in professional research and to see the dramatic changes such experience can have on these teachers.

As public school teachers have found out, it is one thing to teach about a subject, and quite another matter to use the subject consistently in one's teaching behaviors. So too for teacher educators! It is necessary, if we are to teach from a data base, for teacher educators to model what we teach (practice what we preach). Our dedication to the spirit of inquiry can be communicated to our students by our demonstration of the behaviors of inquiry.

This paper reports on a study conducted in a basic educational psychology course which serves as a gateway course to a teacher education program at Michigan State University. The purpose of the study was to focus on the development of new knowledge, while at the same time teaching about the educational research process and the subject being studied. We chose to study a new format for objective testing; namely, alternate-choice questions. In the course we teach evaluation concepts in testing, e.g. fairness, reliability and validity. So we proposed to model the concepts we were teaching by researching the evaluation process as we evaluated the students' performance in the course.

The Setting

The course in which this study was conducted is Teacher Education 200, The Individual and the School. The course is required of all elementary and secondary majors in the standard teacher education program at M.S.U. It is designed to meet the State of Michigan certification requirement that candidates must study "how human beings grow and learn." Therefore, the course is primarily an introduction to educational psychology. In the past the course has been, at various times, a large lecture course, a small section-small group discussion course, and a self-paced mastery-learning course. The present structure of the course has been in operation for two years and is a coordinated, multi-sectioned small group course. This means the course is coordinated as to content, textbook, tests, term paper and supervised teaching interns serving as instructors of several sections. On an average the course has about 150-200 students enrolled in sections of 22-25. The teaching interns meet weekly with the course coordinator and a course consultant to develop common objectives and strategies and to solve common problems. While we value the students having a relative common experience from section to section, we also value the vitality the individual instructor brings to a small group. All lectures, discussions, demonstrates and presentations are the responsiblity of the section instructor. At the same time, the section instructor can call upon the resources of a staff and of professorial supervision to provide support. The TAs are considered teaching interns for the academic year and participate in a common seminar specifically designed for the teaching of this course. At the time of this study (1982-83) the course material was divided into five units of instruction, including a unit on measurement and evaluation.

Goals of the Project

The learning of sound principles of measurement and evaluation is an integral part of teacher preparation. We sought to model an interest in and commitment to effective and theoretically sound measurement and evaluation.

Specifically, the goals of the project were to:

1. conduct a research study into the usefulness of a novel test item format to evaluate student achievement;

2. increase the reliability and validity of the unit examinations and the final examinations in the course;

3. collect information useful in evaluating course effectiveness, student learning, and teacher performance;

4. facilitate student learning of practical and sound measurement theory and techniques through active involvement in a study investigating an important measurement question;

5. and to model the spirit of inquiry, impressing on pre-service teachers the importance and value of experimentation with new techniques and the necessity of continuous evaluation of both new and well-accepted practices.

The Problem

When it came to the unit in TE 200 on measurement and evaluation, the question which faced us was how to model behavior when it is so abstract to our students. In the other four units of the course we were able to develop simulation assignments which paralleled the real world. For example, in the unit on development we asigned a project in which the students had to apply Piaget's demonstrations of the conservation of continuous quantities. However, with the concepts in measurement and evaluation there were restraints which compelled a delayed application of evaluating behaviors. Our students do not grade papers, make up tests, or study real standard tests during their preteaching preparation. We wanted to heighten the immediacy of the evaluation concepts by involving the students in their own evaluation in a research project focused on test formats.

The choice of investigating alternate choice question format had several advantages in response to our problem. First of all, it was an original study. While Ebel (1980)

proposed the new test format, which he called alternate choice questions, no major study had been conducted on reliability or validity of that format of testing compared with multiple choice questions. Secondly, the focus on a testing format permitted the inclusion of a topic already within the course content. Thirdly, we had already identified a need to improve the quality of the testing program in the course. So, finally, the project presented a realistic application of a current and valid research study which had immediate application to the world in which our students lived; namely the fair evaluation of their own achievement in the course.

Procedure

Before describing the procedure that we employed it is appropriate at this point to provide a brief description of alternate-choice (AC) items. Ebel proposed a unique two-choice item format. He called the format alternate-choice items because they are based on a single proposition rather than complex multi-dimensional situations and because they offer only two alternatives instead of the conventional three, four, or five. AC items also differ from multiple-choice (MC) items, including the conventional format for two-choice items, in that they include the responses as segements of a continuous sentence rather than listing them in a column under the stem. For example:

Modeling is a) less b) more effective when verbalization is kept to a minimum.

According to Weiner's theory, feedback that one's efforts on a geography project led to a good grade is likely to produce a) greater b) less effort by highly motivated students.

The crucial variable in operant conditioning is the consequences of the a)stimulus b) response.

Bruner would place a child who uses visual imagery to organize his/her thought at the a) enactive b) iconic stage.

This unique format makes AC items quite easy to write and adaptable to a variety of subjects and testing situations and objectives. Since AC items are based on a single proposition, they tend to be quite short and appear, at least on the surface, to reduce item difficulty (error) associated with item complexity.

In a preliminary study conducted at Michigan State University, Ebel (1980) found that AC items compare quite favorably with true-false items. The results of that study indicate that tests composed of AC items tend to be: 1) easier, 2) more highly discriminating, and 3) considerably more reliable than tests composed of true-false items. Ebel (1980) also noted that students seemed to prefer AC items and perceived them to be less ambiguous than true-false items.

The authors of this paper became aware of AC items in the Fall of 1982. We were intrigued with the claims made for them and elected to include a small number of them in unit exams on an experimental basis. We were pleased with the ease of writing AC items, the item statistics for those experimental items were quite satisfactory, and the items were well-received by the students. It seemed to us that this unique item form had great potential as an important addition to the repertoire of test item forms that teachers have at their disposal, but also, that a need existed for more empirical investigation to evaluate the usefulness of this test item form. At the same time, we were considering the problem of modeling a spirit of inquiry and actively involving pre-service techers in a measurement project. A course research project involving a comparision of the psychometric properties of AC and MC items seemed to address both these needs.

When we proposed such a research project, we were fortunate and pleased to find a great deal of support and encouragement among the course staff and College of Education Faculty. The Office of Evaluation Services agreed to assist in test construction and analysis, including a variety of computer programs for test scoring and item and test analysis. Several professors of educational psychology expressed willingness to serve as technical consultants on test design and construction, item review and editing and research design. The support of the teaching interns was vital. Since they actually taught the course and had direct contact with the students, the project would have been impossible without their willing participation.

With the assistance of these resources, we developed a research design. Specifically, we wanted to compare the mean item difficulty and discrimination, reliability, standard error, and item efficiency (mean time required for students to respond to an item) of tests composed of AC and MC items. Since the students were active participants in the study, student perceptions of AC items were also considered valuable information. We developed the research design shown below to gain the information we desired.

The MC items used were available from a large bank of items used previously in the course. These MC items were the product of several item analyses and revisions and were of high technical quality. Since we did not at that time possess a similar pool of AC items, we were forced to create one. In order to ensure that all items used in the final data collection were technically sound and to provide an opportunity to "iron out" the details, we elected to pilot the study in Winter term of 1983 and, if the pilot proved successful, to perform the full research project in Spring term, 1983.

In writing the AC items we continued to find that it is quite easy compared to MC items. The resources used in writing the AC items were: the course text and teacher's manual, the course notes, manual, and projects, common students questions and problems, ideas from the teaching interns and educational psychology faculty, and the bank of MC test items. As items were written they were reviewed and edited by faculty measurement specialists for psychometric quality and by the teaching interns and educational psychology faculty members for subject matter importance and accuracy. Based on these reviews, items were either retained, revised or discarded. Using this method we were able to develop a bank of approximately 250 AC items.

The next phase in the pilot project involved designing and constructing the tests for Winter term, 1983. Since each unit test and the final exam was to be composed to two content-parallel, equivalent subtests---one subtest composed of AC items and the other composed of MC items—special care had to be taken in test construction. A table of test

TABLE 1

Research Design for Pilot Study
Winter, 1983

Test Analysis: Alternate Choice vs. Multiple Choice

	Mean Item Difficulty	Mean Item Discrimination	Kuder-Rich. Reliability	Standard Error	Number of Questions
Test 1 AC					30
Test 1 MC					30
Test 2 AC					30
Test 2 MC					30
Test 3 AC					30
Test 3 MC					30
Test 4 AC					30
Test 4 MC					30
Final AC					50
Final MC					50
					340

specifications was designed for each test which included unit objectives, cognitive level to be tested, and the number of items to be allocated to measure each objective. These tables of specifications were created with the aid of the teaching interns and educational psychology faculty. The tables allowed us to ensure that the AC and MC tests were measuring the same concepts at the same level of cognitive complexity, and that the AC and MC tests were in fact content parallel and equivalent. Incidently, we made the tables of specifications available to the students before the tests and they reported that the tables were a great aid in studying for the tests.

Methodology

The pilot study performed in Winter term 1983 was successful and we were encouraged by our findings to proceed with the full research study in the Spring Term. Analysis of the data gathered in the pilot study revealed (see Table 2) that as a rule, the mean item difficulty, discrimination and reliability of the AC tests were only slightly lower than the MC tests. In the pilot study we were able to try-out our newly written AC items and on the basis of item analysis and student reaction were able to refine them further. Many of our original items were revised or rewritten, and some we were forced to discard. Also, on the basis of input from the students and teaching interns, we found that four unit tests and a final exam involved an excessive amount of testing time. Consequently, we eliminated a test and administered only three unit tests and a final examination in the Spring Term.

In Spring Term, 1983 we conducted the full study and collected the data for final comparisons. In addition to comparing the psychometric properties of AC and MC item, we were interested in knowledge about the efficiency of AC items. Specifically, we wanted to know the ratio of the number of AC items that students are able to respond to in a given time period compared to MC items. Such information would pertain to the reliability of tests composed of AC items. We, therefore, conducted a time study on the first two unit tests given in the Spring Term, 1983. In order to conduct that time study, it

TABLE 2

TE 200, Winter Term, 1983

Test Analysis: Alternate Choice vs. Multiple Choice

N = 144

	Mean Item Difficulty	Mean Item Discrimination	Kuder-Rich. Reliability	Standard Error	Number of Questions
Test 1 AC	36%	26%	.4791	2.3671	30
Test 1 MC	30%	26%	.5090	2.2350	30
Test 2 AC	17%	24%	.6290	1.9064	30
Test 2 MC	26%	29%	.6500	2.2005	30
Test 3 AC	19%	18%	.2783	1.9538	30
Test 3 MC	28%	28%	.5524	2.2679	30
Test 4 AC	25%	24%	.4866	2.0707	30
Test 4 MC	24%	25%	.5950	2.1318	30
Final AC	25%	21%	.6055	2.7383	50
Final MC	33%	28%	.7503	2.9926	50
					340

was necessary to create two alternate forms for the tests of unit one and two. The alternate forms were identical in content. However, Form A presented the thirty AC items first and the thirty MC items second; and Form B reversed the order. Students in all sections were timed and asked to mark the item they had just completed at the end of five minutes and at the end of ten minutes. The marked answer sheets were then tabulated. We were not surprised with the results. In Ebel's study (1980) he determined that students were able to answer almost exactly the same number of AC items as true-false items, The results of our tabulations indicated that our students were able to answer almost two AC items for each MC item that they attempted. (See Appendix).

Since student perceptions of all of the tests and particularly of AC items were valued as an important consideration in testing, we constructed and administered a survey designed to assess those perceptions. Students were asked to respond frankly and to write additional comments on the form. Care was taken to assure student anonymity in their feedback. Remarks made by the students under the comment section of the survey and verbally to the teaching interns indicated that the students did report honestly and took the task seriously.

Results

The results of the study conducted Spring Term 1983 are summarized in Table 3. Alternate forms generated for the first two unit tests (to facilitate the collection of information on item efficiency) account for the differences between Tests I and II vs. Test III and Final Exam.

It is interesting to note that, in almost all cases, the tests composed of AC items are slightly less difficult and slightly less discriminating than the tests composed of MC items. Correspondingly, when we compare the reliabilities of the AC and MC tests composed of equal numbers of items, the reliability estimate for the AC tests are again, in most cases, slightly less than the reliability estimates of the MC tests. Taking into

TABLE 3

SPRING TERM '83

Test Analysis Alternate Choices vs Multiple Choices

N = 112		Mean - Item Difficulty		Mean - Item Discrimination		Kuder-Rich. Reliability		Standard Error		# of Questions
		A	B	A	B	A	B	A	B	
TEST 1	AC	36	32	28	32	.5604	.7114	2.3536	2.2507	30
(AC Spearman-Brown Corrected)						(.67)	(.80)			
	MC	34	29	27	36	.5405	.7785	2.8246	2.1788	30
		A	B	A	B	A	B	A	B	
TEST 2	AC	28	27	25	27	.4909	.5585	2.2189	2.2257	30
(AC Spearman-Brown Corrected)						(.61)	(.67)			
	MC	31	32	33	30	.6942	.6306	2.3107	2.3216	30
TEST 3										
(AC Spearman-Brown Corrected)						(.60)				
	AC	31		26		.4886		2.3214		30
	MC	28		32		.6825		2.2367		30
	COMBINED	29		25		.7391		3.2328		60
FINAL										
	AC	26		24		.6819		2.7691		50
(AC Spearman-Brown Correc						(.77)				
	MC	35		26		.7168		2.9157		50
	COMBINED	30		22		.8201		4.0241		100

NOTE: Coefficients in parentheses indicate the AC reliability coefficient corrected with the familiar Spearman-Brown formula. It is an estimate of the AC reliability coefficient equate for time.

consideration the fact that the reliability estimates for quite short tests, the reliabilities of all tests are in the satisfactory range.

The results of the time study to determine the efficiency of AC items were as we expected. Previous studies have shown that students usually are able to answer two true-false items for every MC item and we expected similar results from AC items. We, in fact, did find that the students were able to respond to almost twice as many AC items as MC items in a given time period (See Appendix on Time Study) important implications for the reliability estimates of AC tests. Longer tests with more items are invariably more reliable than shorter tests with fewer item, all other things being equal. With the knowledge that students are able to respond to nearly twice as many AC items as MC items in a given time period, it is apparent that for true comparison AC tests should be compared to MC test with testing time held constant. That is, AC tests composed of fifty items should be compared to MC tests composed of thirty items, since students could be expected to respond to either fifty AC items or thirty MC items in the same period of time. We were able to estimate the reliabilities of these theoretically longer AC tests using the familiar Spearman-Brown Correction Formula. The corrected reliability coefficients are included in Table 3 in parentheses. These corected coefficients cast an even more favorable light on AC items. Note that in four of the six cases that the corrected reliability estimated of the AC tests are higher than the reliabilities of the MC tests.

The results of the survey given to the students to assess their perceptions indicated that the students generally viewed the course testing and AC items quite positively. (For summary of Student Survey, see Appendix) Over 70% of the students indicated that the tests provided a strong motivation to learn the principles taught, and that the AC items tested important concepts in the curriculum. While slightly more than half of the students perceived AC items as more difficult and ambiguous, than MC items, those perceptions were not verified by the item statistics. Both the difficulty and discrimination fo the AC items were less than the MC items. From the comments that

the students made and from their responses to the survey, we also learned that a high majority (approximately 75%) of the students felt that AC items are a valuable form and that future course tests should be composed of both AC and MC test items.

In summary, our experiences as researchers of AC items was quite positive. We found AC items much easier to write than MC items and therefore, were able to write many more AC items than MC items in a given time period. We did discover that it is easier to write AC items when aiming for testing lower level cognitive objectives, such as knowledge and comprehension. However, this is also generally true of MC items. Our positive perceptions were shared by our students and were verified by the data showing that AC items compare quite favorably with MC items.

Conclusions and Recommendations

Replications of this study are warranted. While the results were positive, in order to gain more information on the usefulness of this unique item form, AC items should be studied using other subject matter disciplines and with other samples. It would also be of great interest to study the cognitive task involved in responding to AC items and other conventional formats.

We intend to continue using AC items in our testing program. With more practice, we expect to become even more proficient at writing AC items. We wholeheartedly encourage others to try using AC items also. We intend to continue conducting research in TE 200, since all indications are that this project provided a positive learning experience for the students and contributed to their understanding of both research and evaluation in instruction. This learning experience for the students was particularly valuable because the students were directly involved with the process and content of the research. An area of the course's subject matter was illumined in a way that is usually not available in introductory courses such as TE 200.

We recommend that others try writing AC items and using them in student evaluation. You will find that they are easy to write, efficient, and produce good

psychometric results. We would appreciate your informing us of your results using AC items, expecially as you use them to assess students' learning at different cognitive levels. We welcome colleagues in further explorations of AC item use, efficiency and relevance.

BIBLIOGRAPHY

Burke, J. B., The Individual and the School, Course Syllabus and Student Manual for TE 200, East Lansing, Michigan State University, 1983.

Ebel, R. L., Some advantages of alternate-choice test items. Unpublished manuscript, Michigan State University, 1980.

Good, T. L. and Brophy, J. E., Educational Psychology, A Realistic Approach. 2nd Edition, New York, Holt, Rinehart, and Winston, 1980.

APPENDIX A

TE 200
Spring, 1983

TEST AND ITEM QUESTIONNAIRE

The perceptions and impressions of students taking tests are an important criteria in evaluating the tests and items of which they are composed. In the interest of improving the tests used in this course please take a few moments to answer this short questionnaire. Your responses will provide valuable information for future test planning and construction.

When responding to these questions please consider only the tests and test items you have experienced in this course.

Circle the response of your choice.

SA = Stronly Agree
A = Agree
D = Disagree
SD = Strongly disagree

---

1. The tests given in this course provide a strong motivation for me to study. SA A D SD

2. Alternate-choice items tend to be more difficult than multiple-choice items. SA A D SD

3. Most of the alternate-choice items test important points in the curriculum. SA A D SD

4. Many of the alternate-choice items are ambiguous.
SA A D SD

5. Alternate-choice items are challenging and do a good job of testing how much I know. SA A D SD

6. I would prefer taking a test composed of alternate-choice items to a test composed of multiple-choice items. SA A D SD

7. It is usually easier to interpret and understand the question posed in alternate-choice items than in multiple-choice items.
SA A D SD

8. Multiple-choice items tend to be more ambiguous than alternate-choice items. SA A D SD

9. In the future, unit exams should be composed of

   a. all alternate-choice items.
   b. all multiple-choice items.
   c. part alternate-choice items and part multiple-choice items.

10. Please write any comments below.

APPENDIX B

TE 200: Test and Item Questionnarie Summary

Questionnaire Item	Percent Agree/Strongly Agree	Percent Disagree/Strongly Disagree
1. The tests given in this course provide a strong motivation for me to study.	71%	29%
2. Alternate-choice items tend to be more difficult than multiple-choice items.	60%	40%
3. Most of the alternate-choice items test important points in the curriculum.	77%	23%
4. Many of the alternate-choice items are ambiguous.	59%	41%
5. Alternate choice items are challenging and do a good job testing how much I know.	59%	41%
6. I would prefer taking a test composed of alternate-choice items to a test composed of multiple choice items.	34%	66%
7. It is usually easier to interpret and understand the question posed in alternate-choice items than in multiple-choice items.	45%	55%
8 Multiple-choice items tend to be more ambiguous than alternate-choice items.	32%	68%
9. In the future, unit exams should be composed of:		
a. all alternate-choice items	5%	
b. all multiple-choice items	21%	
c. part alternate-choice items and part multiple-choice items.	74%	

APPENDIX C

SUMMARY OF TIME STUDY
TE 200 TESTS I AND II, SPRING, 1983

Question: How many A-C items do students complete as comapred with M-C items, given the same amount of time?

TEST I (N=97)	After 5 Minutes	After 10 Minutes
Mean number of Alternate Choice items completed	15.3	26.6
Mean number of Multiple Choice items completed	10.4	17.4
Differences	4.9	9.2

TEST II (N=110)	After 5 Minutes	After 10 Minutes
Mean number of Alternate Choice items completed	14.4	26.7
Mean Number of Multiple Choice items completed	8.6	16.2
Differences	5.8	10.5

Mean Differences of combined data (Test I & II)	After 5 Minutes	After 10 Minutes
(Number of A.C. items more than M.C. items)	5.35	9.85

Ratio of M.C. to A.C. = 1:1.6
(60% more A.C. than M.C. in equal time)