A STUDY AND CLASSIFICATION OF TEACHER-MADE SCIENCE TESTS IN SECONDARY SCHOOLS OF MONTGOMERY COUNTY USING TAXONOMY OF EDUCATIONAL OBJECTIVES OF THE COGNITIVE DOMAIN AS THE CRITERION OF JUDGMENT

BY

ADNEY ELDRIDGE CROSS, III

A STUDY AND CLASSIFICATION OF TEACHER-MADE SCIENCE TESTS IN SECONDARY SCHOOLS OF MONTGOMERY COUNTY USING TAXONOMY OF EDUCATIONAL OBJECTIVES OF THE COGNITIVE DOMAIN AS THE CRITERION OF JUDGMENT

A Research Paper

Presented to

the Graduate School

Austin Peay State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in Education

---

by

Adney Eldridge Cross, III

June 1972

To the Graduate Council:

I am submitting herewith a Research Paper written by Adney Eldridge Cross, III, entitled "A Study and Classification of Teacher-made Science Tests in Secondary Schools of Montgomery County Using Taxonomy of Educational Objectives of the Cognitive Domain as the Criterion of Judgment." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts in Education.

Major Professor

Accepted for the Council:

Dean of the Graduate School

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

INTRODUCTION

Statement

Success by students on tests is an important criterion for determining what the student has accomplished. Of course, all students do not take the same tests. There are various reasons for this. Much testing in the secondary school is done by the classroom teacher. All teachers do not teach the same topics or units for the same courses. Although at times teachers may present the same topics in a course, each teacher may emphasize certain aspects and see some parts as being more important than others. Because teachers' tests are constructed to evaluate what the teachers feel is important, there will be as many different tests as there are teachers. The tests may be quite similar or very dissimilar depending, of course, upon the teachers involved. Not only might the tests differ by testing different material, but they may differ greatly if the material covered is the same. One test may require a greater difficulty of thinking than another. If there are differences, what are they? How different are they? Is one more difficult than the other? Even if the potential for difference is vast, is there still a large degree of likeness among the tests?

Although biology, chemistry and physics are different courses, they are nevertheless somewhat related since they are all laboratory sciences. Do the tests in these fields differ as to the difficulty of thinking involved? Does one field require a more difficult level of thinking than another?

Importance

This study will indicate the level of thinking required of those science students whose tests were studied. The tables will indicate a similarity or lack of similarity found in the thinking required of biology, chemistry and physics students.

Methodology

In Montgomery County, there are three secondary schools: Clarksville High School, Montgomery Central High School and Northwest High School. Teacher-made tests were obtained from the biology, chemistry and physics teachers from these three schools. No physics tests from one school were collected because none of the tests were made by the teacher. The tests were used during the last six-weeks grading period of the 1970-1971 school year. Each test item was categorized according to the classification in Taxonomy of Educational Objectives of the cognitive domain. Tables were constructed showing what percentage of each category was contained in each test. An average of these percentages was found for all tests given by each teacher, and this was

included in a table. An average of percentages was also found for all biology, chemistry and physics tests and included in a table.

Definitions

Teacher-made tests. Teacher-made tests are those tests which are compiled, constructed and administered by the classroom teacher.

Six-weeks grading period. The six-weeks grading period is the unit of time for which grades are periodically determined. The six-weeks grading period is a period of six weeks or approximately thirty school days.

Limitations

The fact that no previous studies have been found necessarily limits the author of this research in his ability to anticipate problems. A further limitation to the study is its subjectivity, for the classification of the test items rests, ultimately, upon the researcher's own judgment. Any study which involves working with people is not entirely free from biases and other subtleties which hamper objectivity, and it is anticipated that complete cooperation will not be achieved.

The study involves only seven teachers. Four teach biology; one teaches chemistry; and, two teach both chemistry and physics. The small number of teachers involved tends to decrease the validity of the study.

The last six-week grading period of the year was considered, and this may not be representative of the entire year. The tests were constructed on different topics. Some topics require a greater difficulty of thinking than do other topics.

Testing may be the only criterion for evaluation by some teachers while others may rely on many different sources to evaluate the student. This and the fact that teachers grade tests differently tends to decrease the validity of the importance of this study.

Chapter 2

TAXONOMY OF EDUCATIONAL OBJECTIVES

The Taxonomy of Educational Objectives was created "to provide for classification of the goals of our educational system," to aid teachers in building "a range of possible educational goals or outcomes in the cognitive area," to "help one gain a perspective on the emphasis given to certain behaviors by a particular set of educational plans," and to help "specify objectives so that it becomes easier to plan learning experiences and prepare evaluation devices."

The complete taxonomy has "three major parts--the cognitive, the affective, and the psychomotor domains. The cognitive domain . . . includes those objectives which deal with the recall or recognition of knowledge and the development of intellectual abilities and skills." It includes behaviors such as ". . . remembering; reasoning; problem solving; concept formation; and, to a limited extent, creative thinking."

The cognitive domain is composed of six divisions. They are, in order: knowledge, comprehension, application, analysis, synthesis and evaluation. Knowledge is the lowest division, and evaluation is the highest. Each succeeding division adds to, and includes, all the preceding

divisions. For example, application involves aspects of comprehension, and comprehension involves certain knowledge. Table I presents the taxonomic structure as found in Taxonomy of Educational Objectives.

Table I. Taxonomy of Educational Objectives--
Cognitive Domain

1.00 Knowledge
- 1.10 Knowledge of specifics
  - 1.11 Knowledge of terminology
  - 1.12 Knowledge of specific facts
- 1.20 Knowledge of ways and means of dealing with specifics
  - 1.21 Knowledge of conventions
  - 1.22 Knowledge of trends and sequences
  - 1.23 Knowledge of classifications and categories
  - 1.24 Knowledge of criteria
  - 1.25 Knowledge of methodology
- 1.30 Knowledge of the universals and abstractions in a field
  - 1.31 Knowledge of principles and generalizations
  - 1.32 Knowledge of theories and structures

2.00 Comprehension
- 2.10 Translation
- 2.20 Interpretation
- 2.30 Extrapolation

3.00 Application

4.00 Analysis
- 4.10 Analysis of elements
- 4.20 Analysis of relationships
- 4.30 Analysis of organizational principles

5.00 Synthesis
- 5.10 Production of a unique communication
- 5.20 Production of a plan, or proposed set of operations
- 5.30 Derivation of a set of abstract relations

6.00 Evaluation
- 6.10 Judgments in terms of internal evidence
- 6.20 Judgments in terms of external criteria

Chapter 3

REVIEW OF RELATED LITERATURE

The awareness of the Taxonomy of Educational Objectives in evaluation is becoming more prominent. Fry (1963) stated, "recent literature in programming has recognized the potential value that an analysis of programs by the Taxonomy might have in obtaining a detailed analysis of the behaviors to be taught."

Michael (1965) cited two studies "in which a conscientious attempt had been made to duplicate in test construction the objectives of the Taxonomy."

Bensen (1967) has proposed that the Taxonomy be used as a tool to coordinate the objectives of the learning process in the field of industrial arts. This would, in some ways, resolve the feud between the two schools of "basic-skills" learning and "problem-solving" learning.

Tyler (1966) proclaimed the Taxonomy as an efficient tool in the evaluation process because "it makes possible some evaluation of a program without the user having resources of hundreds of subjects and a tremendous budget." It would aid the student. "If programmers were aware of the Taxonomy, they might write frames with higher levels of the taxonomy so students could develop higher intellectual skills which are important in our society."

An entire article by Anderson (1967) was devoted to praise of and pleas for the Taxonomy of Educational Objectives." A beginning of the solution to the problem of communication and, more specifically, of a vocabulary has appeared. It is the Taxonomy of Educational Objectives. It is used by this author as the premise on which a test was constructed for her doctoral research."

"This article is a plea for all science educators to investigate the possibility of adopting the terminology in the Taxonomy of Educational Objectives, or if that does not seem palatable or satisfactory, for a similar classification of terms to be developed by the profession. When this is done, it is believed that great progress can then be made in science education."

However, its use is somewhat limited. Lisonbee (1966) has found that of "twenty-two educational objectives listed by one school district . . . only two are subject to evaluation by testing. Of fifteen objectives listed in a physics curriculum guide on what it was hoped students would achieve, achievement in only three can be measured with tests."

Harris (1963) has said that there was a "lack of any new principles being proposed for the development of achievement tests as reflected, for example, by the absence of any influence of the Taxonomy of Educational Objectives."

Studies concerning the Taxonomy of Educational Objectives and evaluation have been limited. Demchik (1970),

however, stated his feeling concerning present day science education. "I feel that emphasis has been only on the lower cognitive levels of development, i.e., knowledge, comprehension, and some application." He devised a test that would measure transfer and the higher levels of the cognitive processes using the Taxonomy of Educational Objectives as a guide. Six questions from all six divisions comprised the test. He administered this and a traditional test. Both had thirty-six items. On the traditional test the mean score was 28.97, and the range was twelve to thirty-five. On the tests using the Taxonomy of Educational Objectives as a guide the mean was 22.6, and the range was ten to thirty-three. One hundred ten students took both tests. There was a "strong significant difference between the traditional test and [the other]."

Davis and Hunkins (1966) categorized questions from the end of chapters in three fifth grade social studies books. Of a total of seven hundred thirty-two questions only two involved evaluation (6.00), only one involved synthesis (5.00) and none involved analysis (4.00). Thirty items, which represented four percent, concerned application (3.00). Fifty-nine items, or nine percent, dealt with comprehension (2.00). The remaining six hundred forty items, which were eighty-seven percent, were knowledge (1.00) questions. They fostered "little critical thinking."

Tyler (1966) in a study involving two hundred eighty-seven questions, found one hundred thirty-two were knowledge (1.00) and one hundred forty-six were comprehension (2.00).

Lochhead (1964), et. al., in a study of midyear examinations constructed by Virginia high school science teachers, found that seventy-eight percent of approximately fourteen thousand test items dealt with the simplest division of the cognitive domain, knowledge (1.00).

Research has shown that the classification of test items is a somewhat complicated task. The Taxonomy of Educational Objectives has stated, "the task of classifying test exercises is somewhat more complicated than that of classifying educational objectives. Before [one] . . . can classify a particular test exercise, he must know, or at least make some assumptions about, the learning situations which have preceded the test. He must also actually attempt to solve the test problem and note the mental processes he utilizes. [He] should also take into consideration the possibility that the processes used in selecting the correct answer in a recognition form of question may be somewhat different from those used in considering the incorrect alternatives in the same question."

This is echoed by Hedges (1966). "One danger . . . [in constructing test items] is the possibility for a particular test item to assume different levels in the classification system depending on the testee's educational background. Thus, Problem X if given to a physicist might require only simple recall for solution, whereas the same problem presented to a high school student might involve mental processes of a high order."

Lonbard (1965) further stated, "a given question may not represent the same task to all students. What may be an analysis question to one student may be a knowledge question to a student who has read an explanation of the phenomenon."

Tyler (1966) found that two raters classifying three hundred eighty-four items in groups of ten had from sixty-three to ninety-seven percent agreement.

Lonbard (1965) made statements concerning the Taxonomy of Educational Objectives and evaluation. "The fine points of distinction between the categories of the taxonomy are of minor importance. Whether a question is of the analysis or evaluation type is irrelevant if otherwise it is a probing question."

"The difficulty of a question does not necessarily depend on its classification in the taxonomy."

"We are not making value judgments about the categories. That is, an evaluation is not, in and of itself, better than a knowledge question. There are worthwhile and non-worthwhile questions in both categories. Tests should not overload on any one type of objective."

The type item, or formats, should be different. Multiple choice is usually inefficient in evaluating some of the abilities of analysis, synthesis, and evaluation."

Hedges (1966) adds, "the multiple-choice type lends itself more readily to the measurement of the higher mental processes than do most other objective test forms. Staunch supporters of the essay question will find that it is used frequently at levels five and six [Synthesis (5.00) and Evaluation (6.00)]."

Chapter 4

ANALYSIS OF DATA

Seven teachers were involved in the study. The teachers have been numbered: the first four are biology teachers; five is a chemistry teacher; and, six and seven are both chemistry and physics teachers. Table II shows the tests that belong to each teacher. There are ten biology tests, fifteen chemistry tests and ten physics tests.

All ten biology tests (Table III) are in many ways similar. All but test number one have 100% of the test items in the lowest classification, i.e., knowledge (1.00), and 91% of the test items in test number one are contained here also. No tests had items from the second class, comprehension (2.00).

All biology tests had a majority of their test items dealing with knowledge of specifics (1.10). All items from test eight were in this division.

Items involving the knowledge of ways and means of dealing with specifics (1.20) were found in all tests except test one and eight. None of the tests contained items testing knowledge of conventions (1.21) or knowledge of methodology (1.25).

Table II. Teachers and Test Numbers

Teacher	Course	Test Number
1	Biology	1
2	Biology	2, 3
3	Biology	4
4	Biology	5, 6, 7, 8, 9, 10
5	Chemistry	1, 2, 3, 4
6-C	Chemistry	5, 6, 7
7-C	Chemistry	8, 9, 10, 11, 12, 13, 14, 15
6-P	Physics	1, 2
7-P	Physics	3, 4, 5, 6, 7, 8, 9, 10

Table III. Number and Percentage of Biology Test Items

Test No.	No. of test items	Taxonomy of Educational Objectives																	
		1.00	1.10	1.11	1.12	1.20	1.21	1.22	1.23	1.24	1.25	1.30	1.31	1.32	2.00	2.10	2.20	2.30	3.00
1	45	41 91%	35 78%	20 44%	15 33%							6 13%	6 13%						4 9%
2	51	51 100%	39 76%	16 31%	23 45%	12 24%		1 2%	11 22%										
3	44	44 100%	35 80%	15 34%	20 45%	9 20%		3 7%	6 14%										
4	50	50 100%	39 78%	21 42%	18 36%	11 22%			11 22%										
5	37	37 100%	21 57%	7 19%	14 38%	16 43%		1 3%	14 38%	1 3%									
6	35	35 100%	31 89%	19 54%	12 34%	4 11%		1 3%	1 3%	2 6%									
7	35	35 100%	21 60%	16 46%	5 14%	13 37%		3 9%	5 14%	5 14%		1 3%	1 3%						
8	19	19 100%	19 100%	13 68%	6 32%														
9	45	45 100%	24 53%	11 24%	13 29%	21 47%		3 7%	15 33%	3 7%									
10	35	35 100%	29 83%	17 49%	12 34%	6 17%			5 14%	1 3%									

Of the items concerning the knowledge of the universals and abstractions in a field (1.30), all were included in the division of the knowledge of principles and generalizations (1.31). Only test one and seven had items from this category.

As with biology, most test items on the chemistry tests (Table IV) were found in the lowest class, knowledge (1.00). Items of test seven were entirely from the class of application (3.00).

Tests twelve and thirteen had all items dealing with translation (2.10). Tests six, eight, nine, ten, eleven and fourteen dealt only with knowledge of specifics (1.10). No items on any test dealt with knowledge of trends and sequences (1.22), knowledge of methodology (1.25), knowledge of theories and structures (1.32) or interpretation (2.20).

In physics (Table V) there were few items concerning knowledge of ways and means of dealing with specifics (1.20) or knowledge of the universals and abstractions in a field (1.30). In all but test one and two there were no items concerned with comprehension (2.00).

Tests one, two, three, eight and nine had application (2.00) items with items of tests three, eight and nine being exclusively application. Test four, five, six, seven and ten had all items dealing with knowledge of specifics (1.10).

Table IV. Number and Percentage of Chemistry Test Items

Test No.	No. of test items	Taxonomy of Educational Objectives																	
		1.00	1.10	1.11	1.12	1.20	1.21	1.22	1.23	1.24	1.25	1.30	1.31	1.32	2.00	2.10	2.20	2.30	3.00
1	25	18 72%	5 20%		5 20%	11 44%	6 24%		5 20%			2 8%	2 8%						7 28%
2	25	14 56%	6 24%	1 4%	5 20%	7 28%			7 28%			1 4%	1 4%		2 8%			2 8%	9 36%
3	25	23 92%	16 64%		16 64%	7 28%	7 28%												2 8%
4	25	24 96%	22 88%	18 72%	4 16%	2 8%			2 8%										1 4%
5	26	21 81%	17 65%	11 42%	6 23%	4 15%	2 8%			2 8%									5 9%
6	28	20 71%	20 71%	5 18%	15 54%														8 29%
7	5																		5 100%
8	18	18 100%	18 100%	3 17%	15 83%														
9	15	15 100%	15 100%		15 100%														
10	12	8 67%	8 67%	1 8%	7 58%														4 33%
11	15	15 100%	15 100%	5 33%	10 67%														
12	15														15 100%	15 100%			
13	9														9 100%	9 100%			
14	40	40 100%	40 100%	23 58%	17 48%														
15	50	33 66%	33 66%	18 36%	15 30%										17 34%	17 34%			

Table V. Number and Percentage of Physics Test Items

Test No.	No. of test items	Taxonomy of Educational Objectives																	
		1.00	1.10	1.11	1.12	1.20	1.21	1.22	1.23	1.24	1.25	1.30	1.31	1.32	2.00	2.10	2.20	2.30	3.00
1	21	9 43%	8 38%	6 27%	2 10%	1 5%				1 5%					9 43%	4 19%		5 24%	3 14%
2	32	20 63%	16 50%	9 28%	7 22%	3 9%		1 3%		2 6%		1 3%	1 3%		5 16%		4 13%	1 3%	7 22%
3	7																		7 100%
4	20	20 100%	20 100%	15 75%	5 25%														
5	25	25 100%	25 100%	19 76%	6 24%														
6	10	10 100%	10 100%	9 90%	1 10%														
7	20	20 100%	20 100%	8 40%	12 60%														
8	6																		6 100%
9	10																		10 100%
10	20	20 100%	20 100%	13 65%	7 35%														

Table VI shows the teachers (two, three and four) who did not have application (3.00) items on their tests. They were all biology teachers.

All but teachers one, seven-C, and seven-P administered items dealing with knowledge of ways and means of dealing with specifics (1.20). Only four teachers (one, four, five and six-P) had questions dealing with knowledge of the universals and abstractions in a field (1.30). Only teachers five, seven-C and six-P asked comprehension (2.00) questions. No items dealing with knowledge of methodology (1.25) or knowledge of theories and structures (1.32) were given.

Observing Table VII reveals that most (98%) biology test items were concerned with knowledge (1.00). The rest were application (3.00). Two-thirds of the chemistry test items dealt with knowledge (1.00) as did over half of the physics test items. Over three-fourths of the biology test items and over one-half of the chemistry and physics test items were related to the knowledge of specifics (1.10).

Ten percent of the chemistry test items and fifteen percent of the physics test items dealt with comprehension (2.00). There were more application (3.00) items than comprehension (2.00) items with twenty-four percent of the chemistry test items and twenty-eight percent of the physics test items having been this type.

There were no test items on any test dealing with analysis (4.00), synthesis (5.00) or evaluation (6.00).

Table VI. Average Percentages of Each Teacher

Teacher	Taxonomy of Educational Objectives																	
	1.00	1.10	1.11	1.12	1.20	1.21	1.22	1.23	1.24	1.25	1.30	1.31	1.32	2.00	2.10	2.20	2.30	3.00
1	91	78	44	33							13	13						9
2	100	78	33	45	22		5	18										
3	100	78	42	36	22			22										
4	100	74	43	30	26		4	17	6		1	1						
5	79	49	19	30	27	13		14			3	3		2			2	19
6-C	51	45	20	26	5	3			3									49
7-C	67	67	19	48										29	29			4
6-P	53	44	28	16	7		2		6		2	2		30	10	7	14	18
7-P	63	63	43	19														38

Table VII. Average Percentage of Each Course

Course	Taxonomy of Educational Objectives																	
	1.00	1.10	1.11	1.12	1.20	1.21	1.22	1.23	1.24	1.25	1.30	1.31	1.32	2.00	2.10	2.20	2.30	3.00
Biology	98	77	41	36	18		2	14	2		4	4						2
Chemistry	66	54	19	35	11	5		5	1		1	1		10	10			24
Physics	58	54	36	18	4		1		3		1	1		15	5	4	7	28

Chapter 5

SUMMARY

This study has shown the similarity and lack of similarity found in the thinking required of biology, chemistry and physics students using the classification in the Taxonomy of Educational Objectives of the cognitive domain as a criterion of judgment.

The tests given to biology students dealt almost exclusively with Knowledge (1.00) which composed 98% of the test items. The remaining 2% dealt with Application (3.00).

Almost two-thirds (66%) of the chemistry test items were concerned with Knowledge (1.00), while a little less than three-fifths (58%) of the physics test items dealt with this area of the cognitive domain. Ten percent of the chemistry test items and fifteen percent of the physics test items concerned Comprehension (2.00), while almost one-fourth (24% and 28% respectively) of the chemistry and physics items dealt with the higher category of Application (3.00).

While, for example, a certain number of test items may be concerned with Comprehension (2.00), these also must deal with Knowledge (1.00). Likewise, those items concerning Application (3.00) also deal with Comprehension (2.00) and Knowledge (1.00).

Although the dissimilarity of the biology, chemistry, and physics tests is quite noticeable, there is much similarity. The difficulty of thinking required for the items of the chemistry and physics tests was more similar than either that of the biology and chemistry tests or the biology and physics tests.

In all three areas of biology, chemistry, and physics, well over one-half of the test items were Knowledge (1.00) items. There were more items in the group of Application (3.00) than in the Comprehension (2.00) group in all three fields. There were no items in the groups of Analysis (4.00), Synthesis (5.00), or Evaluation (6.00), which are the three higher groups of the Taxonomy.

There were several limitations that hindered or made this study less valid. The classification of test items was dependent upon the researcher's own judgment. Although efforts were made to make the decisions objective, this is a herculean task which cannot ultimately be reached.

The study involved only seven teachers, and since each teacher tests differently, it could easily be assumed that the seven chosen for this study are not representative of all science teachers. Only one six-week grading period of the year was used, and the tests of this period may not have been representative of the tests of the entire year.

The validity of this study would be increased if it involved more teachers and more tests that span a longer period of time. Having several skilled evaluators for the

tests would be much better than having only a few people. Finding correlations between grades made on tests and the difficulty of thinking required for the items of such tests would show how the difficulty of thinking is involved with grades. Finding correlations between the difficulty of thinking required for test items and the mental age of those students making correct responses to such items might give hints as to at what level of thinking students perform effectively. This might possibly give clues as to what students are capable of doing at certain mental ages.

Literature Cited

Anderson, J. S. 1967. A critical need. Sci. Ed. 51: 286-287.

Bensen, J. M. 1967. Do your students know--or memorize? Ind. Arts and Voc. Ed. 56, no. 8:24-26.

Bloom, B. S., et. al., 1956. Taxonomy of educational objectives. David McKay Co., New York.

Davis, O. L., Jr. and F. P. Hunkins. 1966. Textbook questions: what thinking processes do they foster? Pea. J. of Ed. 43: 285-292.

Demchik, M. J. 1970. The classroom test . . . a measure of transfer of learning. Sch. Sci. and Math. 70:693-703.

Fry, E. 1963. Teaching machines and programmed instruction. McGraw-Hill, New York.

Harris, C. W. 1963. Problems in measuring change. Univ. of Wis. Press, Madison, Wisconsin.

Hedges, W. P. 1963. Testing and evaluation for the sciences in the secondary school. Wadsworth Pub. Co. Inc., Belmont, California.

Lisonbee, L. 1966. Testing, what for? Sci. Tea. 33, no. 5:27-29.

Lochhead, I. C., et. al., 1964. An analysis of test items to evaluate the objectives of science teachers in Virginia combined and secondary schools. Div. of Ed. Res. Univ. of Vir. Charlottesville, Virginia.

Lonbard, J. W. 1965. Preparing better classroom tests. J. of the Nat'l. Sci. Tea. Assn. 32, no. 7:25-28.

Michael, W. B. 1965. A short evaluation of the research reviewed in education and psychological testing. Rev. of Ed. Res. 35:92-99.

Tyler, L. L. 1966. The taxonomy of educational objectives: cognitive demain--its use in evaluating programmed instruction. Calif. J. of Ed. Res. 17:26-32.