OpenBU

Boston University Theses & Dissertations

http://open.bu.edu

Dissertations and Theses (pre-1964)

1950

A review of the available objective-type tests used in college science courses since 1938

https://hdl.handle.net/2144/5526 Downloaded from DSpace Repository, DSpace Institution's institutional repository Ed. Thesis Ginsberg, R.H. 1950

Stored

BOSTON UNIVERSITY

SCHOOL OF EDUCATION

Thesis

A REVIEW OF THE AVAILABLE OBJECTIVE-TYPE TESTS USED IN COLLEGE SCIENCE COURSES SINCE 1938

by

Rita H. Ginsberg

B.S., Northeastern University, 1946

submitted in partial fulfillment of the requirements for the degree of Master of Education April, 1950

> Boston University School of Education

First Reader: John G. Read, Associate Professor of Science Education

Second Reader: Henry W. Syer, Assistant Professor of Education

ACKNOWLEDGMENT

The author wishes to thank Dr. John G. Read of the Boston University School of Education who supervised this study, made many helpful suggestions, read, and criticized this manuscript.

TABLE OF CONTENTS

List	st of Tables	Page vii
CHAI	DIFED	VII
OIIA		
I	THE PROBLEM, ITS SCOPE, AND LIMITATIONS	3 1
	Statement of the problem Justification Scope Assumptions Objectives Measured by College Science Restatement of the Problem	1 2 3 Tests 5 7
II	RESEARCH IN BIOLOGY, CHEMISTRY, AND PHY TESTING	SICS 8
	Review of the Research in Biology Testi Review of the Research in Physics Testi Review of the Research in Chemistry Tes	ng 8 ng 13 sting 18
III	TECHNIQUES OF TEST EVALUATION	24
	Criteria lists Rating scales	24 26
IV	TEST PROCUREMENT AND EVALUATION METHOD	33
	Test Procurement Evaluation Method Relevant information Format Content Validity Reliability Administration Scorability Interpretation Conclusions	33 34 34 35 357 88 30 30 30 30 30
۷	REVIEW OF BIOLOGY, CHEMISTRY, AND PHYSI TESTS	CS 41
	Organization of Test Reviews I. Cooperative Science Tests A. Biology Test Reviews 1. Cooperative College Bio Test, 1943	41 43 43 10gy 43

	a serie between					
	CHA	PTER				Page
			Β.	Chemis	try Test Reviews	47
		(±)		1. A	.C.S. Cooperative Biochem-	47
				2 1	C S Cooperative Conoral	=
				2. A	hemistry Test 10.8	50
				3. A	.C.S. Cooperative Organic	54
				C	hemistry Test, 1948	24
				4. A	.C.S. Cooperative Physical	57
				C	hemistry Test, 1946	
. ÷				5. A	.C.S. Cooperative Chemistry	60
				. T.	old analysis,	
				6. Å	.C.S. Cooperative Chemistry	62
	1			T	est in Quantitative Analysis.	0.2
				1	.947	
			C.	Physic	s Test Reviews	65
	1			1. C	ooperative Physics Test for	65
		TT	TInid	U ete het	tes Armed Forces Institute	68
		TT •	Test	ts (USA	FI)	00
			Ā.	Biolog	y Test Reviews	68
				1. E	xamination in Biology,	68
				C	ollege Level, 1945	-
				2. E	alloga Lovel 10/5	12
			В	Chemis	try Test Reviews	71.
			ь.	1. E:	xamination in General	74
				C	hemistry, College Level,	
			0.645	1	944	
			с.	Physic	s Test Reviews	77
		8		1. E	xamination in Physics,	11
					Orrege Lever, 1947	
	VI.	TRENDS	5 IN	TESTIN	G WITH STANDARDIZED	79
		INSTRU	IMENI	S		
		2200.000				70
		Intere		in Coop	Standardized Tests	80
		Presei	li ri	end hv	Author's Survey	00
		Proc	edur	e used	in survey	80
		Tabl	Les t	ased or	n survey results	80
						40
	VII	SUMMAP	α, α	ONCLUS.	LONS, AND RECOMMENDATIONS	04
		Summar	y of	Proce	dures Involved	82
		Revi	Lew c	of the l	Research	82
		Summar	y of	the F	indings	83
		Rese	earch	in Bi	ology Testing	83
		Rese	earch	i in Che	emistry resting	6)

CHAPTER		Page
Research in Physics Testing		83
Standardized Tests		84
Conclusions	10 × 10	84
Recommendations		85
BIBLIOGRAPHY		86

25

APPENDIX-List of test publishers whose tests are reviewed

LIST OF TABLES

а ^с ,

.

TABLE		Page
ı.	RESULTS OF QUERIES ON COLLEGE USE OF OBJECTIVE-TYPE TESTS IN BIOLOGY	80
2.	RESULTS OF QUERIES ON COLLEGE USE OF OBJECTIVE-TYPE TESTS IN CHEMISTRY	81
3.	RESULTS OF QUERIES ON COLLEGE USE OF OBJECTIVE-TYPE TESTS IN PHYSICS	81

CHAPTER I

THE PROBLEM, ITS SCOPE, AND LIMITATIONS

<u>Statement of the problem</u>. The purpose of this study is threefold, namely: to review the research in testing which has been done since 1938 in the fields of chemistry, physics, botany, and zoology as taught in the colleges; to review and evaluate the standardized tests which have been published in these fields since 1938; and to indicate the college trend in regard to the use of objective-type tests in chemistry, physics, botany, and zoology courses.

Justification . Much research has been done in the area of college science testing as has been indicated by the fact that the college physics and chemistry testing programs have been reviewed yearly since 1938. The figures from the 1939-1940 college chemistry testing program² indicate that small colleges use the chemistry tests published by the Educational Testing Service³ more than large colleges. The reason civen for the smaller test-use in the larger colleges is that the instructors devote more time to

- L.S. Foster. "The 1939-1940 College Chemistry Testing Program", Journal of Chemical Education, 18:159-164, April, 1941.
- 2. <u>ibid</u>.

3. ibid.

scientific research than to educational problems.

However, according to Tyler¹ large colleges are showing an interest in tests using objective-type items. He mentions for example, the fact that the Ohio State University Botany and Zoology Department has devised a program of study on tests with the purpose of developing more satisfactory examinations in order to measure the progress of the student.

Since there has been no integration of the research or of the tests available in the sciences mentioned, this paper will act as a central depository containing a review of both the research and of the tests published for courses in college chemistry, physics, botany, and zoology.

<u>Scope</u>. Only research involving testing in college chemistry, physics, botany, and zoology since 1938 will be reviewed. General chemistry, qualitative and quantitative analysis, organic chemistry, biological chemistry, and physical chemistry are included in the term, <u>college</u> <u>chemistry</u>. By <u>college physics</u> is meant heat, light, sound, electricity, optics, and atomic structure. ^Botany and zoology could be included under the general term, biology.

Only standardized tests published in the above fields since 1938 will be used for reviewing purposes.

1. R.W. Tyler. "Some findings in the field of college biology", <u>Science Education</u>, 17:133-142, October,1934. The survey which indicates the trend in college science testing involves the sending of a questionnaire to the chemistry, physics, and biology departments of 48 major **colleges** and universities chosen by random sampling from the 1949 issue of <u>The College Blue Book</u>¹ and the 1948 issue of American Universities and Colleges.²

<u>Assumptions made</u>. It is assumed that in colleges throughout the country there is a trend in science testing similar to that indicated by survey of the chemistry, physics, and biology departments of 48 major colleges and universities.

<u>Procedure</u>. The procedure used in this study is as follows:

1. A brief questionnaire involving yes or no answers was sent to the departments of chemistry, physics, and biology in 48 major colleges and universities in order to determine whether or not objective-type test items were utilized.

2. "The Education Index ", "Journal of Educational Research ", "Review of Educational Research ", "The Journal of Chemical Education ", "School, Science, and

- 1. W.H. Hurt. The College Blue Book 5th ed., R.I.: College Blue Book 1949, 464 p.
- 2. <u>American</u> <u>Universities and Colleges</u>. 5th ed., Wash. D.C.: <u>American</u> Council on Education, 1948, 1054 p.

Mathematics ", " American Journal of Physics ", and other bibliographies and periodicals served as a source for the establishment of a working bibliography.

3. Using the working bibliography as a basis, the material relevant to testing in college science was read, digested, and presented as a review of the research.

4. A list of tests were secured from the Cooperative Test Division¹, from Buros², and from Hildreth³. These were limited to college chemistry, physics, botany, and zoology.

5. The tests were obtained and reviewed according to the following criteria list:

a. Relevant information: name, author, publisher, copyright date, cost, science course, forms, time, and

- 1. Cooperative Test Division of the Educational Testing Service, <u>Cooperative Achievement Tests</u>, N.Y.: Educational Testing Service, 1949, 68 p.
- 2. O. K. Buros. The Mental Measurements Yearbook, Highland Park, N.J.: 1941, 674 p.

The Third Mental Measurements Yearbook, N.J.: Rutgers University Press, 1949, 1047 p.

3. G.H. Hildreth. <u>Bibliography of Mental Tests and Rating</u> <u>Scales</u>, N.Y.: The Psychological Corp., 1939, 295 p.

Supplement, <u>Bibliography of Mental Tests</u> and <u>Rating Scales</u>, N.Y.: The Psychological Corp., 1945, 86 p. statement of purpose.

b. Format: printing, paper, diagrams, etc.

c. Content: number of questions, type of questions, and objectives measured. d. Validity: curricular or statistical.

e. Reliability: coefficient of reliability.

f. Administration: time of test, directions for giving, materials used.

g. Scorability: hand scoring by strip key, cutout stencil, or transparent stencil; self scoring; and machine scoring.

h. Interpretation: availability of sufficient norms.

i. Conclusions: summary of test characteristics and data not included elsewhere.

OBJECTIVES MEASURED BY COLLEGE SCIENCE TESTS

In order to review a test the author has to decide which objective or objectives the test is measuring. Robertson¹ states that there are three levels of chemistry education, namely:

- "1. Learning chemical facts.
 - 2. Understanding of the reasoning presented by the instructor or textbook and recitation of such reasoning back in an examination.

1. G. R. Robertson. " Examinations in the practice of elementary organic chemistry ", Journal of Chemical Education, 25:104-105, February, 1948. 3. Selecting data not suggested by the instructor and using such data to solve a problem. "

According to Hendricks and Smith¹ the measurable objectives for college chemistry are:

- " 1. An appreciation of the relationship of chemical knowledge to man's physical, social, and economic life.
 - 2. A knowledge of basic chemical facts, laws, principles and theories.
 - 3. The ability to use the vocabulary, symbols, and formulas of chemistry.
 - 4. The ability to apply the facts, laws, and theories of chemistry.
 - 5. The ability to form generalizations or to set up hypotheses from experimental data.
 - 6. The ability to formulate procedures for testing the reliability of generalizations.
 - 7. The ability to manipulate simple chemical laboratory equipment. "

Tyler² lists 4 objectives to be measured in either botany

or zoology, namely:

- " 1. The ability to recall important facts, principles, and technical terms.
 - 2. The ability to formulate reasonable generalizations from the specific data of an experiment.
 - 3. The ability to plan an experiment which might be used to test a given hypothesis in botany or zoology.
 - 4. The ability to apply general principles to new situations. "

According to the author the following learning outcomes are those which are covered in college science tests:

1. B.C. Hendricks and O.M. Smith. "Measurable objectives for general college chemistry ", <u>School, Science</u>, and <u>Mathematics</u>, 36: 747-752, October, 1936.

2. R.W. Tyler. op. cit.

1. Recall of facts, principles, and concepts.

2. Application of facts, principles, and concepts.

3. Application of problem-solving skills.

4. Application of the scientific method.

The sequence of the scientific method is as follows:

- a. Accurate observation of phenomena.
- b. Formulation of hypothesis.
- c. Testing of hypothesis by experiment.
- d. Statement of theory (a clarified hypothesis).
- e. Further testing by means of varied, repeated, new experiments including those done by many individuals.
- f. If every result over a reasonable time fits the theory, a"law" may be worded.
- g. Classification under "laws" of all new facts, and repeated testing of correct placement by experiments. There is no " law" of physical science which is not subject to change if facts are found which make it incomplete or false.

5. Application of scientific attitudes.

The above list of objectives will be used in reviewing the college chemistry, physics, botany, and zoology tests.

RESTATEMENT OF THE PROBLEM

This study will review the research since 1938 in the fields of college chemistry, physics, botany, and zoology testing. Tests published in these fields since 1938 will be reviewed and the trend in present science testing will be indicated.

1. J.G. Read. "The Scientific Method ", Unpublished teaching material. Boston University School of Education, 1947.

CHAPTER II

RESEARCH IN BIOLOGY, CHEMISTRY, AND PHYSICS TESTING An examination is considered adequate if it measures student progress and achievement. By means of this measurement of growth, the instructor can guide the pupil toward the desired outcomes in science. "Examination questions like personal excellences reveal their qualities by their outcomes. "¹ The following review of the research in college science testing will indicate, where possible, the outcomes the author proposes to measure and the types of items by which these outcomes are evaluated.

In a later chapter the work of the Educational Testing Service and their Cooperative Tests will be discussed and reviewed according to the past and present trends in testing.

RESEARCH IN BIOLOGY TESTING

A test devised to measure certain outcomes of an orientation course in biological science was written by Fleming.² The measuring instrument was designed to evaluate growth of college students in recall of specific

^{1.} B.C. Hendricks. " Examination practice in general college chemistry ", Journal of Chemical Education, 21:85-86, February, 1944.

M.C. Fleming. "An evaluation of outcomes of science in higher education ", <u>Science Education</u>, 27:81-85, November, 1943.

information, understanding of generalizations, elements of problem-solving, and in scientific attitudes. The items for the instrument were selected in the following manner:

- A large number of items was administered to a heterogeneous group of students with no regard to to major field or previous science instruction.
- 2. From the results the items were selected on the basis of internal validity and index of difficulty.
- Items for measuring scientific attitudes were selected from five attitude scales developed by others.
- 4. Finally, the instrument was revised after repeated administration of the test items.

The reliability coefficient using the split-half technique was .83. The test was deemed valid according to the selection of items on a basis of (1) internal validity, (2) the judgment of a professional jury, (3) the response on an attitude scale of advanced graduate students, and (4) the careful selection of factual information. The instrument was administered to an experimental group of 148 undergraduate students who took an orientation course in biological science and to a control group of 80 students who did not take any science course during the experiment. Fleming¹ found that growth which was statistically

1. ibid.

significant was indicated in each outcome measured. Most growth appeared in recall of facts $(8.2 \neq .59)$, and in understanding of generalizations $(4.3 \neq .62)$.

Another test devised to measure student progress in biological science was devised by Mason.¹ He proposed to measure the increase in retention of factual information in the second and third terms work in biology by the comparison of scores on a pre-test with those obtained from a post-test. The test covered the following areas:

- 32 questions on reproduction and heredity
- 20 questions on coordination and adjustment
- 14 questions on evolution
- 18 questions on parasitism
- 6 questions on interrelationships
- 10 questions on economically destructive and useful

plants and animals,

The mean for the pre-test given to 1171 students was 21.62 and the standard deviation was 10.27. The post-test mean was 47.36 and the standard deviation was 12.84 for a group of 482 students. Thus, there was an increase of 25.74 in the mean score indicating that there was an increase in the retention of facts. The greatest growth

J.M. Mason. " Pre-test and post-test study in biological science", Journal of Educational Research, 42:228-233, November, 1948.

in fact retention was noted in the areas of reproduction and heredity, and the least progress was made in the field of interrelationships.

An instrument was designed by Barnes and Mouser¹ to determine students' beliefs about biological phenomena. From a knowledge of the existing misconceptions in students' minds the authors attempted to develop a critical attitude. A test consisting of 35 items on health, disease, and nutrition; 25 items on behavior of animals; and 13 items on notions relating to psychic phenomena was administered to 65 freshmen students. Each item had a choice of answers, namely: true, partially true, partially false, and false, with relative scores of 3,2,1,0 ranging from true to false. Therefore, the higher the score, the more credulous the student. The test was considered reliable according to a reliability coefficient of .92 obtained by using the Spearman-Brown formula.

Semans² constructed a true-false test with a rebutting feature which is used in all his biology classes. The test is given, marked, passed back, and discussed in the

^{1.} M.W. Barnes and G.W. Mouser. " A comparative performance of high school and university freshmen on a test of biological misconceptions ", <u>School</u>, <u>Science</u>, <u>and</u> <u>Mathematics</u>, 43: 447-450, May, 1943.

^{2.} F.M. Semans. "The Rebuttal Test ", School and Society, 53:419, March, 1949.

following class period. The rebutting feature has the following characteristics:

- If a student by rebutting in essay form can show that he was not guessing in answering incorrectly, and yet the validity of his answer is questionable, he receives one-half credit for his answer.
- 2. If the reasoning is sound, but based on misinformation, he receives three-fourths credit for his answer.
- 3. If the answer is satisfactory, he receives full credit.

A circulating library containing test items was suggested by Hendricks.¹ He devised a plan whereby the best examination questions from many colleges were collected, validated, assembled, and classified. These items were available for teachers in service. Dr. Magnus **01son**² followed this plan and assembled a file of indexed zoology questions. He noted that the second and third use of the item does not alter its validity.

1. B.C. Hendricks. "Better new examinations from old ", Journal of Chemical Education, 17:583-586, December, 1940.

2. ibid.

RESEARCH IN PHYSICS TESTING

The objectives measured by an instrument designed by Bluh¹ are listed as follows:

1. Appreciation of historical situations in science

2. Ability in reading a formula

3. Ability in reading a graphical representation

4. Ability in following a train of ideas

5. Comprehension of a passage of scientific exposition The following is a sample question of the type which measures the ability to follow a train of ideas:

Similar items have been devised to evaluate the other objectives which are listed above.

Schilling³ has designed a " stripped problems test "

1. O. Bluh. " Physics exam and the new curriculum ", American Journal of Physics, 16:20-24, January, 1948.

2. ibid.

3. H.K. Schilling. "Stripped problems test ", <u>American</u> Journal of <u>Physics</u>, 9:124-125, April, 1941.

which evaluates fundamentals in physics. A stripped problem is one which is (1) nearly all divested of mathematics, (2) deals with elementary, concrete situations, and (3) requires physical thinking exclusively, to achieve the solution. The author states that the test is of great diagnostic value and is of great use in the study of units. The test consists of 11 stripped problems similar to the following:

- " An object on a frictionless table has an acceleration of 2 cm./sec.". The horizontal force upon it is 1 dyne. Find the mass of the object. " 1
- " An unbreakable cord of negligible mass is hung over a pulley which is frictionless and has no inertia. A body of mass 1 g is hung from one end, one of mass 2 g is hung from the other end. "ind the acceleration of each body in a region where the gravitational acceleration is 1 cm./sec.² ".

Rassweiler³ evaluates pupils knowledge of physics principles, laws, and definitions with a four part test at the end of each quarter. Part A consists of 30-40 questions on every principle introduced during the quarter. The student has a choice of 10 of the above items. In part B the student has a choice of 10 items from 25-30 simple problems which involve only one principle.

- 1. H.K. Schilling. op. cit.
- 2. ibid.
- 3. Rassweiler. "Improved type of examinations for physics courses ", <u>American Lournal of Physics</u>, 11:223-225, August, 1943.

In part C the student has a choice of 2 problems from 8-12 problems which involve the combination and application of several principles. Part D consists of 4-6 bonus problems which require a thorough understanding of the interrelationship of principles. The student can choose any of these items for extra credit or for the purpose of covering any mistakes made in parts ^A, B, and C. The time for the administration of the exam is 2 hours.

Dunning¹ constructed and evaluated a scientific thinking test in first year college physics. His instrument was designed with the following assumptions in view:

- 1. There are a number of component and independent abilities involved in scientific thinking.
- 2. These abilities may be related to those measured by intelligence and by factual information tests.
- 3. These abilities can be measured by papre and pencil tests constructed in objective form.

Test items were selected according to criteria based on areas of first year college physics, namely: mechanics, heat, light, sound, and electricity. Item validity was established by a pre-test given to 135 students. All items

1. G.M. Dunning. "The construction and validation of a test to measure certain aspects of scientific thinking in the area of first year college physics ", <u>Science</u> <u>Education</u>, 33:221-235, April, 1949.

not reaching an estimated biserial r of .15 and not successfully completed by 15-90 per cent of the examinees were discarded. Finally, the test items were divided into two parts-interpretation of data, and application of principles. The split-half reliability coefficients for the two parts were .96 and .88 respectively. The final test form was administered to 109 freshmen students in a college physics class. Each of these students was also rated by instructors for his ability to interpret data and to apply principles. The students were also subjected to the Strong Vocational Interest Blank and the Thurstone Primary Abilities Test. The following are Dunning's conclusions:¹

- 1. The test was valid and reliable.
- 2. The aspects rated by the teachers correlated to a substantial degree with the aspects measured by the scientific thinking test.
- Interest or lack of interest in science, as indicated by the Strong Vocational Interest Blank, did not influence the scores on the scientific thinking test.
- 4. Factors other than reading ability accounted to a great extent for the differentiation in scores on

1. ibid.

the test.

- 5. The aspects of intelligence measured by the Thurstone Primary Mental Abilities Test did not correlate with the aspects measured by the scientific thinking test.
- According to the statistics a high degree of knowledge of factual material did not indicate a high degree of knowledge of problem-solving.
- 7. There is a moderate positive relationship between interpretation of data and application of principles as indicated by the intercorrelation of scores which was .48/.08.

A test devised to measure achievement in the physics laboratory was devised by Weber.¹ Designated areas in the laboratory are equipped with the material necessary for the solution of an assigned problem. Brief answers are written on a paper beside the number which corresponds to the number on the designated area. Student technique is observed and graded by the instructor. The following are examples of problems from Weber's laboratory test:²

- 1. "<u>Vernier caliper</u> Vernier caliper; meter stick; piece of wire. Measure diameter of wire to 2 significant figures. "
- 2. " Friction Board of medium length; block; ruler. Determine the coefficient of static friction."

 L.R. Weber. "Dynamic tests for the laboratory ", <u>American Journal of Physics</u>, 12:101-103, April, 1944.
ibid.

RESEARCH IN CHEMISTRY TESTING

The following review of the research in chemistry testing will indicate that there is an increase in the number of tests which measure laboratory outcomes as compared with the tests designed for the laboratory in college biology and physics. Tests which eliminate guessing and which measure applications of historical experiments will also be included in the research review.

An historical experiment test constructed by Wellings¹ consists of a statement of an experiment performed in the past by a famous chemist with the addition of one or more questions related to the experiment. The following is an example of a typical test item:

- " In 1667 John Mayow inverted a large glass globe over a lighted candle standing in water. The water rose inside the globe. When the flame died out a large bulk of gas was left.
 - 1. Diagram the experiment at its conclusion.
 - 2. Diagram this experiment at the instant the globe inverted over the candle touched the water.
 - 3. Why did the water rise inside the globe?
 - 4. Was anything added to the gas left in the flask?
 - 5. Why did the flame die out?
 - 6. Did any chemical change take place in the experiment?"

The test has many advantages, namely:

1. Guessing is eliminated.

2. The application of factual knowledge to an experiment-

1. Wellings. " A new type of test in chemistry ", <u>School</u>, <u>Science</u>, <u>and Mathematics</u>, 39: 351-353, April, 1939.

2

al problem is of value.

3. Exercise of reasoning ability is necessary.

4. Appreciation of great scientists is aroused.

5. Self-expression is possible.

Angell¹ eliminated paper and pencil tests in freshman chemistry courses by developing a punchboard which consisted of 29 items testing knowledge of facts and principles, 45 nonmathematical applicational problems, and 26 mathematical applicational problems. The same test was administered in paper form to a different group. The reliability coefficient obtained by both the split-half method and the Spearman Brown Formula was .93. The conclusions arrived at were:

- The difference between scores on the punchboard group and the paper and pencil group favored the punchboard group.
- 2. The punchboard group made more gain on applicational items than on factual knowledge.
- Students who used the punchboard wanted to continue its use in other exams.
- 4. Students viewed the punchboard as an opportunity

^{1.} G.W. Angell. "Effect of immediate knowledge of quiz results on final examination scores in freshman chemistry ", Journal of Educational Research, 42:391-394, January, 1949.

for learning chemistry, since they knew if they were right or wrong in their answers immediately-a red light glowed for the correct answer.

An examination designed to distinguish between "laboratory thinkers" and "cookbook slaves" has been constructed by Robertson.¹ He states that there are three levels of chemistry education, namely:

- " 1. Learning chemical facts.
 - 2. Understanding of the reasoning presented by the instructor or textbook and recitation of such reasoning back in an examination.
 - 3. Selecting data not suggested by the instructor and using such data to solve a problem." 2

He attempts to measure the third level by stating a problem and by allowing the students one hour to solve the problem with the use of any book which may be of help.

Wright² and Ashford⁴ devised tests which eliminate guessing. The test designed by Wright consists of both a

1. G.R. Robertson. op. cit.

2. ibid.

- 3. W.A.E. Wright. "The modified true-false item applied to testing in chemistry ", <u>School</u>, <u>Science</u>, <u>and</u> <u>Mathemat-</u> ics, 44:637-639, October, 1944.
- 4. T.A. Ashford. "Objective test items of the recognition type that test reasoning and minimize guessing ", Journal of Chemical Education, 19:86-89, February, 1942.

statement which is either true or false and a correction of the statement if it is false. Q is the symbol for false, and <u>-</u> is the symbol for true. An example of a test item is as follows:

"Column I Column II Silicon occurs in nature in 0 combined the free state. "1

The word or words which make the statement true or false are underlined and are corrected if false.

Ashford's² test uses paired items which require students to make a prediction and to give the reason for the prediction. Guessing is eliminated by writing decoys which are apparently more plausible than the correct responses and by giving credit only when both the prediction and reason for the prediction are correct. The following is one of the paired items:

" The uranium atom has an atomic weight of ______ and an atomic number of _____. The number of protons of which the nucleus of the uranium atom is considered to be built up is: _____54; ____92; ____146; ____238; ______30; and the number of neutrons is: ______54; _____92; ____146; ____238; ____330. " 3

Laboratory tests devised in chemistry are of two types: (1) paper and pencil tests and (2) practical tests which require the actual use of apparatus.

1. W.A.E. Wright. <u>op</u>. <u>cit</u>. 2. T.A. Ashford. <u>op</u>. <u>cit</u>. 3. <u>ibid</u>. From an unpublished list of objectives Hendricks¹ constructed a paper and pencil instrument for the laboratory. He mentioned that even though paper and pencil tests may have only tolerable validity, they will help to determine what the chemistry teaching program is doing. An example of a test item is as follows: 22

"The volume of the liquid as read upon the buret diagrammed below is: 1. 30.0 ml. - 35 -

A test which measured laboratory skills requiring the use of apparatus was designed by Boeck.³ Students were assigned to stations in the laboratory. At the station was the apparatus used in the test situation and a question and answer sheet. The student was alloted l_2^1 minutes to analyze

- 1. B. C. Hendricks. " Paper and pencil tests for the laboratory ", <u>Journal of Chemical Education</u>, 22:543-546, November, 1945.
- 2. ibid.

2. 29.0 ml. 3. 25.4 ml. 4. 29.5 ml. " 2

3. ^C.H. Boeck. "A practical examination of skills and techniques acquired in freshman chemistry ", <u>Science</u> <u>Education</u>, 31:329-324, December, 1947. the situation and then was required to move on to the next station. The examination was administered to 40 students in a 50 minute class period. The following is **an** example of a question on measurement:

" Situation: A triple beam balance is balanced with a salt placed directly on the pan. Question: The Material on the pan is balanced. Indicate the total weight and what is wrong with the set-up. "1

The following conclusions were reached:

The exam seemed desirable because the student was:

- 1. confronted with situations similar to actual working conditions.
- 2. required to make an actual observation.

3. not given as much opportunity to guess.

4. required to use greater discriminating power.

Adams² rated his general chemistry laboratory students on manual skill, manipulative technique, efficient use of laboratory time, comprehension of the problem, self reliance, and orderliness. During a year he recorded 32 graphical ratings on a line from low to high.

1. ibid.

2. C.S. Adams. "The importance of laboratory work in general chemistry at the college level ", Journal of Chemical Education, 19:266-270, June, 1942.

CHAPTER III

TECHNIQUES OF TEST EVALUATION

Objective-type or standardized tests are evaluated in two ways. The tests reviewed in this study will be evaluated by a criteria list. The author will establish a list of the important characteristics of a good examination and will review each test for these characteristics. The other and less flexible method of test evaluation is the use of rating scales. In this case point values are assigned to various features of the test, but since different individuals will not agree in their point scores, the author prefers the first method, namely: evaluation by a criteria list. In this chapter will be found both accepted rating scales and accepted criteria lists.

CRITERIA LISTS

Hoff lists seven criteria of a properly constructed standardized test, namely:

- 1. The test must be valid. 2. The test must be reliable.
- 3. The test should care for individual differences.
- L. The routine of scoring should be cut down to a minimum.
- 5. The items should be as brief as possible.
- 6. The examination should be improved and revised frequently.
- 7. Directions should be complete and simple.
- 1. A.G. Hoff, Secondary School Science Teaching, Philadelphia .: The Blakiston Company, 1947, pp. 195-198.

Dubins¹ lists as his criteria of a good test the following:

1. Pertinent information

- a. name of test
- b. copyright date
- c. publisher
- d. distributor
- e. author's name
- f. cost
- g. number of forms
- h. time
- i. purpose
- 2. Description
 - a. content
 - 1. memorization of facts
 - 2. ability to apply fundamental concepts and principles
 - 3. the use of the scientific method
 - 4. attitudes
 - b. appearance
- 3. Validity
- 4. Reliability
- 5. Administration
- 6. Scoring
- 7. Interpretation
- 8. Reviewer's conclusions

The distinguishing characteristics of a good examination

as discussed by Greene² are listed below:

- 1. Validity
- 2. Reliability
- 3. Adequacy
- 4. Objectivity
- 5. Administrability

 M.I. Dubins. " A review of the research in testing in secondary chemistry and physics from 1938-1948 including 14 reviews of standardized tests ", unpublished Master's Thesis, Boston: Boston University School of Education, 1949.

2. H. A. Greene, A.N. Jorgensen, and J.R. Gerberich. <u>Measurement and Evaluation in the Secondary School</u>, New York.: Longmans, Green and Co., 1943, pp. 52-74.

6. Scorability 7. Comparability 8. Economy 9. Utility Criteria similar to those of Dubins are presented by Haynes.2 1. Basic facts a. name of author or authors b. publisher of tests c. cost d. copyright date e. number of forms f. time g. grades h. purpose 2. Description a. content-number and type of questions b. physical appearance-paper, print, diagrams, etc. 3. Validity 4. Reliability 5. Administration 6. Scoring 7. Interpretation 8. Conclusions-material not falling in other categories The author's list of criteria which have been stated briefly in the first chapter will be presented in detail in the following chapter. The second method used in test evaluation will be listed below. RATING SCALES The Cole-von Borgersrode' rating scale which is most 1. M.I. Dubins. op. cit.

- J.H. Haynes. "An evaluated list of standardized tests in mathematics for grades 9-12 ", unpublished Master's Thesis, Boston: Boston University School of Education, 1948.
- 3. R.D. Cole and F. von Borgersrode. " A scale for rating standardized tests ", <u>School of Education Record</u>, University of North Dakota, <u>14:11-15</u>, <u>1928</u>.

complete is presented here: Cole-von Borgersrode Scale for Rating Standardized Tests I. Preliminary Information 1. Exact name of test 2. Name and position of author 3. Name of publisher and nearest address 4. Cost 5. Date of copyright 6. Purpose of test II. Validity 25 A. Curricular 15 1. Exact field or range of education functions which test measures. 2. Ages and grades for which intended 3. Criteria with which material was correlated 4. Do questions parallel good teaching procedures? 5. How wide is sampling of topics? 6. What is social utility of questions? 7. Is test claimed to be diagnostic? B. Statistical 10 1. Correlated against what outside criteria? 2. Size and coefficient of correlation 3. Size and representativeness of sampling 4. Proof of validity of items III.Reliability 25 A. Most important items 1. orrelated with what? 2. Size and representativeness of sampling 3. Reliability coefficient 4. The means of the distribution 5. The standard deviations of the distributions 6. If measure other than above is given to prove reliability, what is it? 7. Inter-correlations B. Less important but desirable 1. Order of giving various forms of test 2. Is test reliable for individual measurement? 3. Evenness of scaling 4. Are pupils accustomed to this type of test? IV. Ease of Administration 15 1. Manual of directions 2. Simplicity of administration 8 3. Alternate forms 3 L. Time needed for giving 1 V. Ease of Scoring 10

1. Degree of objectivity 2. Are adequate directions given? 3. Is scoring key adjusted to size of test? 4. Time needed to score one test 5. Simplicity of procedure VI. Ease of interpretation 20 1. Norms 6 2. Is class record provided? 3. Are ther provisions for graphing results? 4. Is interpretation of raw scores easy or hard? 5. Application of results 10 VII. Miscellaneous 5 1. Typography and make-up 2. Is the time required for giving as small as is consistent with reliable measurement? 3. Is the cost in keeping with the amount, scope, and reliability of the results yielded? 4. Is good test service provided by the publisher? 5. Kind of new-type questions used Ross has revised the point system of the Cole-von Borgersrode² scale and has included items IV, V, VI, and VII under one heading. His revised scale is as follows: I. Preliminary information II. Validity 50 A. Curricular 30 B. Statistical 20 III.Reliability 20 A. More important items 15 B. Less important items 5 IV. Usability 30 A. Ease of administration 10 B. Ease of scoring 5 . Ease of interpretation 10 D. Miscellaneous 5 1. C.C. Ross. Measurement in Todays Schools, New York .: Prentice Hall Inc., 1947, p. 192. 2. R.D. Cole and F. von Borgersrode. op. cit.
A check list for briefing and evaluating tests has been constructed by Rinsland.¹ The following is a reproduction of this new type rating scale:

Author(s)	Name of test	Publisher	Date
Check	Item		Description
	I. KIND		
1.1	ACHIEVEMENT - EDUC	CATIONAL	
1.	Name of subject		
2.	Test		
3.	Scale of quality		
4.	Battery of subject	ts	
5.	Development, think	king	
6.	Evaluations		
1.2	MENTAL OR INTELLI	GENCE	
1.	Individual		
2.	Group		
1.3	APTITUDE (SPECIFY	ABILITY)	
1.	Special subjects		
2.	Mechanical-manual	-motor	
3.	Clerical-commerci	al	
4.	Professional		
5.	Artistic		
1.4	PERSONALITY AND C	ONDUCT	
1.	Ratings		
2.	Adjustment questi	onnaire	
3.	Interest question	naire	
4.	Attitude question	naire	
5.	Tests of conduct,	knowledge, jud,	gment
6.	Performance		
7.	Free association		10 at 11
8.	Physiological mea	sure of emotion	S
1.5	FACTORIAL QUESTIO	NNALKES	q
1.6	SUBJECTIVE TESTS,	NOT EVALUATION	5
1.7	PHYSICAL ASPECTS	TTOC	
1.8	EVALUATION OF OBJ	P OD TECHTUE POD	1.40
	II. SUBJECTIVE O	R OBJECTIVE FOR	C1VI
1. N	ame of objective I	orms	
2. 5	ubjective or evalu	acton form	
- 0	111. USES 10-5		
1. General survey			
2• ±	Togress		
a second eveluating			
1. H.D. Rinsland. " A form for prieting and crainabing			

standardized tests ", Journal of Educational Research, 42:371-375, January, 1949.

29

3. Placement 4. Diagnostic 5. Instructional 6. Clinical 7. Evaluations 8. Guidance and counseling 9. Research only IV. BY WHOM USED 5-2 1. Survey officer 2. Administrator 3. Supervisor L. Teacher 5. Clinician 6. Counselor 7. Researcher V. GROUPS DESIGNED TO MEASURE 5-2 1. Grade level: Pre-school to college 2. Age groups 3. Other groups VI. MECHANICAL DIVISIONS OF TESTS 1. Name and no. of comparable forms 2. Number of parts 3. Number of sittings VII. LENGTH 1. Total time-giving Time of parts
Overall time, whole test
VIII. VALIDITY 25-12 8.1 CURRICULAR 1. Textbook 2. Course of study 3. National committee report 4. Survey of extrinsic or social usage 5. Survey of errors 6. Pooled judgment 7. Experimental analyses 8.2 STATISTICAL 1. Correlation between entire test and : r 2 : r : r 3 2. Item criteria PSYCHOLOGICAL 8.3 IX. RELIABILITY 15-7 1. Half with half corrected by Brown-Spearman 2. Equivalent forms 3. Kuder-Richardson 4. One form with itself

31

5. S.D. or P.E. of distribution or 1,2,3, or 4 6. Standard error or probable error of a score 7. Ratio of reliability (Item 6 divided by item 5) 8. Number and range of population 1. Grades and no. of cases 2. Age and no. of cases 3. Other groups X. NORMS 10-5 1. Grade or grade and month 2. Age 3. Percentile 4. T score, R score, Army score, z score, etc. 5. Scaled score 6. Letter grades 7. Population for norms XI. SCORING DEVICES 5-2 1. Printed key 2. Punched stencil 3. Machine IBM 4. By whom scored 5. Scale of quality INTERPRETATIVE DEVICES 10-5 XII. 1. Profile: percentile, age, or grade 2. Percentile graph 3. Diagnostic record by details 4. Follow up for teaching 5. Reference to texts 6. Other records or devices COST XIII. 1. Tests per 25 2. Manual 3. Diagnostic or other class records 4. Answer sheets 5. Keys 6. Any other materials XIV. CRITICISM 10-5 1. Validity Authority 2. Reliability Authority 3. Comments 5-2 XV. FINAL RATING 100-50 1. Ratings of whole test

The following is a reproduction of a rating scale designed by Otis.

1. Manual 2. Validity 20 3. Reliability 10 4. Ease of administration 20 a. Little special preparation 4 b. Adequate detailed directions 6 c. Time limits clearly stated 6 d. Alternate forms available 4 5. Reputation 6. Ease of scoring 15 a. Objectivity 8 b. Convenient form of key 4 c. Time required 3 7. Ease of interpretation 20 a. Types of norms 10 b. Directions for 3 c. Class record sheet 2 d. Remedial program 5 Typography and makeup 5 8. Total 100 Since the recent trend in test evaluation is the use of

Since the recent trend in test evaluation is the use of criteria lists, the college science tests listed in this study will be evaluated by a list of criteria similar to that of Dubins.²

1. A.S. Otis. "Scale for rating tests ", Test Service Bulletin # 13, Yonkers, N.Y.: World Book Co., 1926

2. M.I. Dubins. op.cit.

CHAPTER IV

TEST PROCUREMENT AND EVALUATION METHOD

The following chapter will explain the procedures by which the standardized tests were procured and the procedures by which the standardized tests were evaluated.

TEST PROCUREMENT

From sources such as the test bibliographies of Buros¹ and Hildreth² and the test library at the Boston University School of Education it was noted that all recent college science tests were published by the Cooperative Test Division of the Educational Testing Service.³ A test catalogue was ordered and specimen sets of the most recent forms of tests in college chemistry, physics, botany, and zoology were obtained.

EVALUATION METHOD

In the past chapter examples of methods of test evaluation such as criteria lists and rating scales including a highly specialized check list were presented. The author has chosen the following criteria list as an effective and complete way to evaluate a standardized test:

- 1. O.K. Buros. op. cit.
- 2. G.H. Hildreth op. cit.
- 3. Cooperative Test Division. op. cit.

relevant information, format, content, validity, reliability, administration, scorability, interpretation, and conclusions.

The meaning if these characteristics of a good examination and the way in which they will be used in evaluation tests will be explained in the rest of this chapter.

a. <u>Relevant information</u>. This category describes general facts which are necessary for the purpose of ordering the examinations. These general facts and their meaning as used in test evaluation are as follows:

 name-indicates subject and range for which the test is suitable.

2. author-indicates name or names of author or authors.

- 3. publisher-indicates nearest address of publisher
- 4. copyright date-specifies recent or older test editions
- cost-specifies speciment set cost, bulk quantity cost, manual cost, and score sheet cost.
- 6. forms-indicates presence of similar arrangements of the same test and is of use in determining reliability.
- 7. time- means working time or time required for actual work on test, not for administration of test.
- 8. statement of purpose-specifies exactly what author intends to measure.

b. Format. The general arrangement of the paper in regard to clarity of print, grade of paper, individuality

in diagrams, and spacing of answers is described under the heading-format.

c. <u>Content</u>. The outcomes measured by the tests, the types of test items, and the number of items measuring each outcome fall in the above category. The objectives measured by multiple choice, recall matching, true and false, and other types of alternate response items should be as follows:

1. recall of facts, principles, and concepts

2. application of fact, principles and concepts

3. application of problem-solving skills

4. application of the scientific method

5. application of scientific attitudes

d. <u>Validity</u>. Among the numerous ways of determining validity the most common are face validity, curricular validity, and statistical validity.

According to Rulon¹ the most conclusive proof that a test is valid is the fact that it has face validity. If questions such as the following can be answered about the test, it is considered to have face validity:

1. Is this the subject matter we are teaching our students

to understand?

1. P.J. Rulon. "Validity of Educational Tests ", Test Service Notebook #3, N.Y.: World Book Co., 1947, p. 1.

2. Is this the way we are trying to get them to understand the subject matter?

The second method used in test validation is curricular validation. If a test deals with the types of educational outcomes desired by the teacher, and if it is at the proper level of difficulty for the student, Greene¹ considers it as having curricular validity. The guiding sources for determining curricular validity are textbooks, courses of study, reports of national or regional committees, and writings of subject specialists. In general, on achievement tests the teacher is the best judge of validity, since he knows what was covered in class.

The next method of validating tests is a more objective type involving statistical techniques. The basic idea involved in statistical validity is a correlation between test scores and ratings by teachers, expert judges, and other standardized tests. If the correlation is high the test is deemed statistically valid. The test is valid if (1) test scores and teacher's marks are closely related, (2) if test scores and judgment by experts are closely related, and (3) if test scores on one standardized test are closely related to test scores on another standardized test which purports to measure equal outcomes.

In the evaluation of the tests in the next chapter

1. H.A. Greene. op. cit. p.55

the type of validation for each test will be discussed.

e. <u>Reliability</u>. ^Greene¹ defines reliability as " the degree to which a test measures what it does measure." He states that a test cannot be valid unless it is reliable. Rulon² uses an illustration to show that a test with little validity may be very reliable.

"It has been shown that the speed of the chirping of crickets varies with the temperature of the air in which they chirp. A blind person could thus, get a better temperature observation by listening to crickets than he could by trying to see the height of a column of mercury."

Reliability may be determined by the following methods:

- Reliability coefficient. Two equivalent forms of the same test are given successively to the same group of students and the scores are correlated.
- Retesting coefficient. One test is given to the same group of students twice under the same conditions and the scores are correlated.
- 3. Chance-half coefficient. One test is given and the scores on odd numbered items are correlated with the scores on even numbered items.

If the correlations are high in all the above methods

1. H. A. Greene. <u>op. cit</u>. p.61 2. P.J. Rulon. <u>op. cit</u>. p.3

the test are reliable. A coefficient of reliability of .87 indicates a reliable test.1

In the review of college science tests the method of determining reliability will be indicated if the information accompanies the test.

f. Administration. This category must be considered from both the teacher and the student viewpoint. In the analysis of test administration by Greene² a test manual should contain directions to be followed verbatim by the teacher and the test should be specific in providing the user with statements on the following items:

- 1. Number of sub-parts in the test.
- 2. Directions for each part of the test.
- 3. Fore exercises to acquaint the pupils with the
- methods of response. 4. Directions for procedure at the bottom of each page and at the end of each test part.
- 5. Definite statements of time limits.
- 6. Definite statements of total possible scores on each test part.

The manual should also specify the materials to be used and those materials which must be provided by the students. The above characteristics of a well-administered test will be used as criteria by the author in evaluating the administration of college science test.

g. Scorability. There are two ways of scoring tests-

1. M.I. Dubins. op. cit. p.34

2. H.A. Greene. op. cit. p.68

by hand and by machine. Hand scoring methods consist of the following devices:

- 1. Strip keys. The correct answer is placed on a strip of cardboard in such a way that the spacing corresponds to the spacing of the test answers. The key is placed alonside the test answers.
- 2. Cutout stencils. A strip of paper is provided with openings which frame the correct answer. This type of stencil is used in correcting tests with scattered answers.

Machine scoring methods consist of a special answer sheet marked by graphite pencils. The correct answer is recorded by the machine by an electrical impulse sent from the graphite answer to the machine.

The type of scoring key provided with each test will be indicated in the test reviews.

h. <u>Interpretation</u>. By the interpretation of a test is meant the comparison of results on the test. In order to interpret results three types of norms are used:

- Grade norms. These are tables of values representing average performance for students in different grade levels.
- 2. Age norms. These are tables of values representing average performance for students in different age groups.

3. Percentile norms. These are tables of values representing either " (1) the percentage of pupils exceeding each score or each number of equally spaced scores, or (2) the score below which a certain percentage of pupils fall. "¹

The procurement of norms, size of norm groups, and status of norm groups will be included in the test reviews.

i. <u>Conclusions</u>. A summary of the characteristics of the tests and the inclusion of material not covered by the other categories will be presented under the above topic.

1. H.A. Greene. op. cit. p.88

CHAPTER V

REVIEW OF BIOLOGY, CHEMISTRY, AND PHYSICS TESTS ORGANIZATION OF TEST REVIEWS

This chapter which consists of the reviews of 12 college science tests is organized in the following manner:

1. The tests are divided into two main categories, namely: (1) test published under the direction of the Cooperative Test Service, and (2) tests published under the direction of the United States Armed Forces Institute (USAFI).

2. Each main category is further divided into three sections: (1) Biology Test Reviews, (2) Chemistry Test Reviews, and (3) Physics Test Reviews. The tests are arranged alphabetically in each of the three sections. The following is an outline of the test organization:

I. Cooperative Science Tests

A. Biology Test Reviews

Cooperative College Biology Test, 1943
B. Chemistry Test Reviews

- 1. A.C.S. Cooperative Biochemistry Test, 1947
- A.C.S. Cooperative General Chemistry Test, 1948
- A.C.S. Cooperative Organic Chemistry Test, 1948

- 4. A.C.S. Cooperative Physical Chemistry Test, 1946
- 5. A.C.S. Cooperative Chemistry Test in Qualitative Analysis, 1948
- 6. A.C.S. Cooperative Chemistry Test in Quantitative Analysis, 1947

C. Physics Test Reviews

1. Cooperative Physics Test for College Students, 1939

II. United States Armed Forces Institute Tests (USAFI)

A. Biology Test Reviews

1. Examination in Biology, College Level, 1945

2. Examination in Botany, College Level, 1945

B. Chemistry Test Reviews

1. Examination in General Chemistry, College Level, 1944

C. Physics Test Reviews

1. Examination in Ohysics, College Level, 1945

COOPERATIVE SCIENCE TESTS

BIOLOGY TEST REVIEWS

I-A-1 Relevant information. name-COOPERATIVE COLLEGE BIOLOGY TEST; author-C. Pearson et. al.; publisher-Cooperative Test Service; copyright date-1943; cost-25 test booklets, 2.50, 25 answer sheets, .90, scoring stencils, set, .30, specimen set, .25; forms-provisional form T; time-90 minutes; purpose-to be used as an end of year achievement test or placement test which measures the objectives of the usual first-year course in college biology and includes items on general structures and functions of plants and animals; more specific items on plants and animals; items on man as representative of mammals; and items on the development of living organisms.

Format. The print is clear but too small. The grade of paper is of good heavy quality, and the answer spaces are well placed. The test has 12 diagrams which are well drawn but are on a too small a scale for clarity. For example, in the heart diagram it is very difficult to distinguish between the arrow pointing to the right auricle and the arrow pointing to the auriculo-ventricular valve.

<u>Content</u>. The test consists of a total of 176 multiple choice and matching type items. It is divided into the following three parts:

1. Part 1-70 multiple choice questions on structures and functions of plants and animals.

2. Part 2-58 items to be matched on plant and animals and the development of living organisms.

3. Part 3-48 items to be matched based on structure and the function of the structure which is designated by $\mathbf{43}$

a diagram.

The only outcomes considered in this test are recall of facts, principles, and concepts, and application of facts, principles, and concepts. The items measuring the above outcomes are scattered throughout the test; the test parts are not divided in such a way as to measure the individual outcomes.

<u>Validity and Reliability</u>. The Cooperative Science Tests do not mention specific reliability coefficients and specific validity indices in the test manuals. However, the catalogue¹ accompanying the tests describes in outline form the manner in which tests are constructed so as to insure the validity and reliability. The outline is as follows:

- a. Preliminary planning and selection of content Analyses of curricula, textbooks, research studies ^Hormulation of objectives and determination of general plan Preparation of detailed test outlines based upon survey of materials Submission of outlines to authorities for criticism Revision of test outlines in accordance with suggestions of critics
- b. Preparation and editing of test items Writing of items by editors and cooperating experts Submission of items to authorities for criticism Revision of items in view of suggestions received Preparation of experimental forms of test
 - c. Administration of experimental forms to a representative sampling of students to obtain item difficulty and validity indices and to detect items which may

1. Cooperative Test Division of the Educational Testing Service. op. cit. be weak or ambiguous

d. Preparation of final forms Selection and revision of items for tentative final form Obtaining from experts in subject-matter fields, test technicians, etc., suggestions and criticisms of the tentative final form Revision and final editing of the test based on the criticisms and suggestions received

e. Administration of final form of test with earlier forms for equating and determination of scaled scores (reliability coefficients) and percentile norms

It can be assumed from this outline that all the tests of the Cooperative Test Division are curricularly and statistically valid and sufficiently reliable.

<u>Administration</u>. The introductory sheet of the test provides students with the necessary general directions. Each of the 3 parts of the test are provided with specific directions for procedure. The test manual contains the directions for teacher use which are complete and can be used verbatim.

The total time limit for the test is 90 minutes. This working time is divided among the 3 parts as follows:

Part 1 40 minutes for 70 items

Part 2 30 minutes for 58 items

Part 3 20 minutes for 48 items

The accompanying manual specifies that the test user should have two No. 2 pencils and an eraser.

The directions for the test would be more efficient if explanatory exercises preceded each of the 3 parts. The time limit for the test parts seems to be adequate, since the student is not expected to finish any one part. The directions on the test booklet and in the test manual are clear and concise.

<u>Scorability</u>. The test can be scored by hand or by machine. Strip keys are provided for hand scoring and the IBM answer sheet is provided for machine scoring. Scoring directions are listed in the last section of the test booklet and in the accompanying manual.

<u>Interpretation</u>. Percentile norms accompany the test. These are tentative end-of-year norms for college biology students based on 667 cases from returns of 16 colleges.

<u>Conclusions</u>. A biology test should include an equal number of items on plants and animals, if it is to be used as an en-of-year achievement test. In this test the lack of items on plants is noticeable. For example, in Part I approximately 45 items deal with animals and 25 with plants. The test should be revised to include more items on plants and to measure additional outcomes other than recall of facts and application of facts.

CHEMISTRY TEST REVIEWS

I-B-1 Relevant information. name-A.C.S. COOPERATIVE BIOCHEMISTRY TEST; author-G.H. Pritham et. al.; publisher-Cooperative Test Service; copyright date-1947; cost-25 test booklets, 2.25, 25 answer sheets, .80, scoring stencils, each.15, specimen set, .25; forms-form x; time-100 minutes; purpose-to be used either as a semester examination which measures fundamental topics in biological chemistry or as an end-of-year achievement test in biological chemistry in college or in medical school.

Format. The test is printed on a good heavy grade paper. The print of the test items is of good size and clear. Two graphs are included in the test-both well-constructed. All answers are placed on a separate answer sheet.

<u>Content</u>. The test consists of a total of 134 5 part multiple choice items. It is composed of the following parts:

1. Part 1-75 items covering fundamental topics included in a one semester course in biological chemistry.

 2. Part 2-59 items covering advanced topics included in a year course in biological chemistry. Scattered throughout the test are items which measure (1)recall of facts,
(2) application of facts, (3) application of problemsolving skills, and (4) application of the scientific method.

Validity and Reliability. See section I-A-1, Validity and Reliability.

Administration. General directions are provided for the student on the cover sheet of the examination. Introductory

directions precede each of the two test parts. Directions for the examiner are listed on both the cover sheet and the test manual. Both student and examiner directions are clear and concise.

The total time limit is 100 minutes. This includes 50 minutes working time on 75 items in part 1 and 50 minutes working time on 59 items in part 2.

Material for the test mentioned in the manual includes either two No. 2 pencils and an eraser, if the test is to be hand-scored, or a special mechanical pencil, if the test is to be machine-scored.

Although the directions are clear, an explanatory exercise preceding the test would be of value. This test could be used in parts-part1 as an end-of-semester achievement test in fundamental aspects of biological chemistry, and parts 1 and 2 as an end-of-year achievement test.

<u>Scorability</u>. The test can be scored easily by hand or by machine. A special answer sheet is provided which can be corrected by a cut-out stencil or by an electrical scoring machine. The scoring formula is provided in the test manual and on the cut-out stencil.

<u>Interpretation</u>. Percentile norms covering Part 1 and Part 2 and the total test are provided. These norms are based on 239 cases from returns of 4 colleges. <u>Conclusions</u>. The test seems to be adequate. It is enhanced by the fact that the parts can be administered to different groups. However a more inclusive set of norms covering more colleges would be of value in the interpretation of the examination.

I-B-2 Relevant information. name-A.C.S. COOPERATIVE GENERAL CHEMISTRY TEST FOR COLLEGE STUDENTS; author-C.E. Ronneberg et. al.; piblisher-Cooperative Test Service; copyright date-1948; cost-25 test booklets, 2.50, 25 answer sheets, .80, scoring stencils, each, .15, specimen set, .25; forms-form 1948; time-110 minutes; purposeto measure general knowledge and information in the field, application of principles, scientific method, and laboratory technique.

Format. The paper is of good quality, and the print is large and clear. An advantage in this test is the fact that the important sections of the test items are emphasized by heavier print than that found in the less important sections. The test has one well-constructed graph and 14 excellent drawings of laboratory apparatus. All answers are placed on a separate answer sheet.

<u>Content</u>. The test contains 135 5 answer multiple choice items. The test items measure the following outcomes:

Part I-30 items which measure general knowledge and

information.

Part II-40 items which measure the application of principles.

Part III-20 items which measure the quantitative application of principles i.e. problem-solving skills.

Part IV-30 items which apply to the scientific method. Part V-15 items which measure laboratory technique. The test covers the range of facts, principles, and concepts of an elementary college chemistry class in an adequate way and has a short section on atomic energy.

The section covering laboratory technique can be given as a separate test. As a paper and pencil test on laboratory methods, part V can be considered adequate, although some of the items test recall of information instead of laboratory technique.

Validity and Reliability. See section I-A-1, Validity and Reliability.

<u>Administration</u>. General directions are privided for the student on both the cover sheet of the test and as an intriduction to each test part. An explanatory exercise is also included in the beginning of the test. Directions for the teacher are included on the cover sheet and in the test manual. The directions are as clear as possible with the added inclusion of the explanatory exercise.

^The total working time is 110 minutes with the addition of the laboratory technique section and 103 minutes excluding this section. ^The working time is divided in the following manner:

Part I General knowledge and information-30items-12 minutes Part II Application of principles-40 items-40 minutes Part III Quantitative application of principles-20 items-

27 minutes

Part IV Scientific method-30 items-24 minutes Part V Laboratory technique-15 items-7 minutes

> Boston University School of Education

Material for the test mentioned in the test manual includes either No. 2 pencils and an eraser for hand-scored tests or mechanical pencils for machine-scored tests.

The time limits for parts I, II, III, and IV seem to be adequate. However, the working time for the laboratory technique section could be increased. The test might be of more value if the laboratory technique section were administered at a time other than the time for the administration of parts I, II, III, and IV.

<u>Scorability</u>. The test is designed for both hand-scoring and machine-scoring. A cut-out stencil is provided for hand-scoring, and the special answer sheet can be run through an electric scoring machine if mechanical pencils are used. Scoring directions are listed in the manual and on the scoring key.

<u>Interpretation</u>. Percentile norms for the individual parts are provided. Norms for the total of parts I,II, III, and IV and for the **to**tal of parts I, II, III, IV, and V are provided. These totals norms are of value since part V is optional and need not be included in the test. The college norms are based on 1639 cases drawn systematically from returns of 8381 cases. All cases had completed two semesters of study in general college chemistry.

<u>Conclusions</u>. The test seems to serve its purpose as an end-of-year achievement **test**. However, it is questionable as to whether part V actually measures laboratory techniques. Many of the items in this section seem to measure recall of facts or problem-solving skills and not laboratory methods. I-B-3 Relevant information. name-A.C.S. COOPERATIVE ORGANIC CHEMISTRY TEST; author-O.C. Dermer, E.F. Degering, et. al.; publisher-Cooperative Test Service; copyright date-1948; cost-25 test booklets, 2.50, 25 answer sheets, .80, scoring stencils, each. .15, specimen set, .25; formsform w (revised 1948); time 100 minutes; purpose-to provide a measure of achievement for use in college organic chemistry classes.

Format. The paper is of good quality, but the print is too small. No diagrams are included in the test and the answers are placed on a separate answer sheet.

<u>Content</u>. The test consists of a total of 90 5 answer multiple choice questions. It is divided into two parts which measure the following outcomes:

1. Part I-45 items covering general information, application of principles, and problem-solving. This section of the test is devised to measure achievement of students who have completed the first semester course in organic chemistry.

2. Part II-45 items composed of three sections which cover (1) general information, (2) application of principles, and (3) problem-solving. The achievement of students who have finished two semesters of organic chemistry is measured by this section.

The test is a well-balanced measuring device in that it covers nomenclature, isomerism, proof of structure, stepwise syntheses, and identification problems. It can be used as an end-of -year achievement test or as an end-ofsemester achievement test.

Validity and Reliability. See section I-A-1, Validity and Reliability.

<u>Administration</u>. Complete student directions are listed (1) on the cover sheet, (2) on each of the two test parts, and (3) in fore exercises in each of the sections of the two parts. Teacher directions ar explained in the test manual and on the cover sheet.

The total working time of 100 minutes is divided in the following manner:

Part I Total-50 minutes

Section A-15 minutes-15 items

Section B-15 minutes-15 items

Section C-20 minutes-15 items

Part II Total-50 minutes

Section D-15 minutes-15 items

Section E-15 minutes-15 items

Section F-20 minutes-15 items

Material necessary for the test includes either two No. 2 pencils and an eraser, or special mechanical pencils, if the test is to be machine-scored.

<u>Scorability</u>. Two mwthods for scoring are indicated: (1) hand-scoring with a cit-out stencil, and (2) machinescoring with an electrical scoring machine, ^Detailed scoring instructions are on the last page of the test. <u>Interpretation</u>. Percentile norms are provided. Each section included in part I has separate norms indicated. Each section included in part II has separate norms. Also, included are separate norms for the total of part I and II. The norms are based on data for 1039 students from 25 colleges and universities who had completed two semesters of study in organic chemistry.

<u>Conclusions</u>. This test is particularly valuable, because it can be used as an end-of-semester achievement test as well as an end-of-year achievement test. I-B-4 Relevant information. name-A.C.S. COOPERATIVE PHYSICAL CHEMISTRY TEST; author-D.A. Dreisbach et. al.; publisher-Cooperative Test Service; copyright date-1946; cost-25 test booklets, 2.50, 25 answer sheets, .80, scoring stencils, each, .15, specimen set, .25; forms-form w; time-100 minutes; purpose-to measure achievement at the end of a physical chemistry course of two semesters.

<u>Format</u>. The test is printed on good quality paper, but the print is too small. Nine figures are included in the test. 8 of these figures are graphs, and 1 consists of 4 apparatus set-ups in electrochemistry. All the figures are constructed on too small a scale for ease of interpretation. A special answer sheet is provided with the test.

<u>Content</u>. The test consists of 65 5 answer multiple choice items which are divided into the following parts:

Part 1-13 items on gases, liquids, solids, and atomic structure

Part 2-12 items on solutions and colloids

Part 3-14 items on homogeneous equilibrium, heterogeneous equilibrium, and kinetics

Part 4-12 items on thermodynamics and thermochemistry Part 5-14 items on electrochemistry

All parts of the test contain items which measure general information, application of information, and problem-solving skills. The test is an adequate measuring instrument, since it includes the important phases of physical chemistry.

Validity and Reliability. See section I-A-1, Validity

and Reliability.

<u>Administration</u>. Directions are provided for student use on both the cover sheet of the exam. and in the introduction to each part. Teacher directions are included on the cover sheet and on a separate directio sheet. No test manual accompanies the test. The student directions are concise but those for the teacher are completely inadequate. Although the total working time of 100 minutes is provided, ther is no mention of time allowance for each part of the test. ^Ther is also no mention of the type of equipment necessary for the test.

The administration directions for the test are extremely poor and hinder the possibility of actually giving the test to a group.

<u>Scorability</u>. Complete scoring directions are provided on the last page of the test, on the direction sheet, and on the cut-out stencil. The test can be machine or handscored by using either the cut-out stencil or an electrical scoring machine.

<u>Interpretation</u>. ^College norms-percentile norms-are provided for each of the 5 parts of the test and for the total test. These are based on returns of 319 cases from 22 colleges. All students were tested at the end of 2 semesters of study.

Conclusions. It is necessary that a proper administration

manual be provided in order to use the test. A more inclusive set of norms should also be provided for more adequate interpretation. I-B-5 Relevant information. name-A.C.S. COOPERATIVE CHEMISTRY TEST IN QUALITATIVE ANALYSIS; author-R.D. Reed, et. al.; publisher-Cooperative Test Service; copyright date-1948; cost-25 test booklets, 2.00, 25 answer sheets, .80, scoring stencils, each, .15, specimen set, .25; forms-form Y; time-100 minutes; purpose-to measure achievement in one-semester courses in qualitative analysis.

<u>format</u>. The paper is of good heavy quality, but the print is too small. No diagrams are included in the test and answers are placed on a special answer sheet.

<u>Content</u>. The total of 80 5 answer multiple choice items is divided among the following parts:

Part I Descriptive Information 45 items

Part II Problems and Equations 15 items

Part III Chemical Equilibrium 20 items The test is a well balanced measuring device, since it covers (1) application of the laws of solution, (2) equations of typical analytical reactions, and (3) problems dealing with solubility product, ionization constant, pH, and common-ion effect.

Validity and Reliability. See section I-A-1, Validity and Reliability.

<u>Administration</u>. Complete student directions are provided by the general directions on the cover sheet, the introductory explanations preceding each part, and the fore exercises included in each part. The test manual and the cover sheet contain examiner directions.

The working time of the test is 100 minutes. The time

limit is divided in the following way:

Part I-20 minutes-45 items

Part II-40 minutes-15 items

Part III-40 minutes-20 items

Part I requires less working time than the other parts, because it measures recall of information, whereas Parts II and III require the solving of problems.

The manual states that the examinees should be provided with two No. 2 pencils and an eraser or with a special mechanical pencil if the test is to be machine-scored.

<u>Scorability</u>. A special answer sheet is utilized which can be hand or machine-scored. A cut-out stencil is provided for hand-scoring. ^Complete scoring directions are printed on the cut-out stencil.

<u>Interpretation</u>. Individual percentile ranks are listed for each part of the test and for the total of the three parts. These ranks are based on data for 738 students from 12 colleges and universities who have received instruction in a one semester qualitative analysis course.

<u>Conclusions</u>. The test seems to present a comprehensive review of a qualitative analysis course. The only criticism the author has is that the print should be enlarged.

I-B-6 Relevant information. name-A.C.S. COOPERATIVE CHEMISTRY TEST IN QUANTITATIVE ANALYSIS; author-R.L. Van Peursem et. al.; publisher-Cooperative Test Service; copyright date-1947; cost-25 test booklets, 2.50, 25 answer sheets, .80, scoring stencils, each, .15, specimen set, .25; forms-forms X and Y; time-ll0 minutes; purposeto measure the achievement of students who have had one or more semesters of quantitative analysis, regardless of whether they have had both gravimetric and titrimetric (volumetric) methods.

Format. The test is printed clearly on heavy grade paper. A special answer sheet is provided. No provisions for answers are on the test sheet. One test item involves a series of 5 graphs which are well constructed.

<u>Content</u>. An outstanding feature in the make-up of the test is the reference sheet to be utilized in solving problems. The reference sheet includes (1) molecular weights, (2) atomic weights, (3) gravimetric factors, (4) logarithms of gravimetric factors, and (5) logarithms of numbers to the base 10.

The test is composed of two parts: (1) 60 items on theory, and (2) 30 problem items. Of the 60 items comprising part 1 23 measure recall of information, 11 measure the application of the scientific method, 9 cover application of information, 9 measure laboratory technique, 7 measure aspects of problem solving, and 1 deals with scientific attitudes. Part 2 consists of the typical calculations concerned with volumetric and gravimetric procedures.

Validity and Reliability. See I-A-1, Validity and Reliability.

<u>Administration</u>. Verbatim directions are provided for the examiner in the test manual and general directions on the cover sheet of the examination. These general directions specify that the student is to be provided with scratch paper and is permitted to use tables of logarithms or a slide rule. Concise student directions which include explanatory exercises are listed before each test part and more general directions are included on the cover sheet.

The total working time is 110 minutes. 35 minutes are alloted to the 60 items on theory and 75 minutes are alloted to the 30 problem items.

Supplies for the test include a slide rule, two No. 2 pencils and an eraser or a special mechanical pencil if the test is to be machine-scored.

It seems that a revision of the time allotment is necessary, since neither part seems to have a sufficient time provision.

<u>Scorability</u>. The test can be hand or machine scored. Cut-out stencils are provided for the marking of the special answer sheet by hand. Scoring directions are listed on the cut-out stencil.

<u>Interpretation</u>. Percentile norms are provided. ^Separate norms are mentioned for part 1 and part 2 and the total test. ^These are based on results for 128 students in 7 colleges and universities. A more complete set of norms for more comprehensive interpretation should be made up.

<u>Conclusions</u>. A more comprehensive set of norms should be provided with the test and it seems that the time allotment should be revised.
PHYSICS TEST REVIEWS

I-C-1 Relevant information. name-COOPERATIVE PHYSICS TEST FOR COLLEGE STUDENTS; author-H.W. Farwell et. al.; publisher-Cooperative Test Service; copyright date-1939; cost-25 test booklets, 3.50, specimen set, .25; formsforms C, D, E, and F; time-215 minutes; purpose-to measure achievement in the following elementary college physics courses: mechanics, heat, sound, light, electricity, and modern physics.

Format. The test is printed clearly on fair quality paper. Well placed answer spaces are provided on the test sheet. No special answer sheets are required. The 12 diagrams included in the various parts of the test are well constructed and concise.

<u>Content</u>. This test is outstanding in its composition because it includes 6 sections which measure achievement in each of the following topics: mechanics, heat, sound, light, electricity, and modern physics. The items are all 5 answer multiple choice type. The following table was prepared by Dubins¹ in an analysis of the sections of the test:

" Questions in Different Groups of Subject Matter in the Different Parts of the Examination M# A QA Total

Modern				
Physics	19			19
Heat	13	7	8	28
Mechanics	14	8	28	50
Sound	3	5	8	16
Light	16	12	7	35
Electricity	15	13	16	44
Total	80	45	67	192
#M	stands	for	retention	of facts;
A	means a	appli	ication of	principles;

1. M.I. Dubins. op. cit. p.63

65

QA means quantitative application of principles." <u>Validity and Reliability</u>. See section I-A-1, Validity and Reliability.

<u>Administration</u>. Directions for students precede each of the sections of the test. Although these directions are concise, explanatory exercises would be helpful. ^Examiner directions are discussed in the accompanying test manual.

The total working time for the test is 215 minutes. This time is divided among the sections in the following way:

Mechanics 50 minutes

Heat 30 minutes

Sound 20 minutes

Light 40 minutes

Electricity 50 minutes

Modern Physics 25 minutes

Each of the above sections may be administered as an individual test.

Materials necessary for the test are two No. 2 pencils and an eraser.

<u>Scorability</u>. Strip keys are provided for the scoring of the test by hand. Scoring directions are explained at the end of each section and on the strip keys. If possible, a separate cut-out stencil should be made by the examiner and the students should be required to place their answers on a separate answer sheet. Thus, the scoring time would be cut to minimum.

<u>Interpretation</u>. National percentile norms are provided for the following:

- 1. Male college students of general physics
- 2. Women college student of general physics
- 3. College physics students enrolled in engineering courses.

Pre-test and post-test results of students who either did or did not study high school physics are summarized in six graphs. Results in mechanics, electricity, heat, light, sound and modern physics are recorded on each of the six graphs.

<u>Conclusions</u>. The test is clear and comprehensive. However, the scoring device should be revised, and a laboratory section should be included. UNITED STATES ARMED FORCES INSTITUTE TESTS (USAFI) BIOLOGY TEST REVIEWS

II-A-1 <u>Relevant information.</u> name-EXAMINATION IN BIOLOGY-COLLEGE LEVEL; author-Examinations Staff for the United States Armed Forces Institute; publisher-American Council on Education; distributor-Cooperative Test Service and Science Research Associates; copyright date-1945; cost-25 test booklets, 2.00, 25 answer sheets, .40, scoring stencil, each, .15, specimen set, .25; forms-form CBg-1-B-4; time-120 minutes; purpose-to measure achievement in the field of college biology.

<u>Format</u>. The test is well printed on good quality paper. A special answer sheet is provided for hand or machinescoring. The six diagrams of the amoeba, hydra, leaf crosssection, vertebrate embryo, vertebrate eye, and bread mold are well constructed. The test also includes two well cut graphs.

<u>Content</u>. The test contains 145 5 answer multiple choice items. It is a well-balanced measuring tool because it covers the following outcomes:

Selection and discrimination of factual material-70 items Identification of structure and functions-30 items

Interpretation of data-45 items

The section measuring the student's ability to interpret data involves a problem presented in paragraph form, tabular form, or graph form from which the examinee draws a conclusion.

The test items are equally divided between animal and plant biology.

Validity and Reliability. No specific validity indices or reliability coefficients are presented in the test manuals accompanying the USAFI tests. In order to insure validity and reliability in the tests , the following method was used in test construction:

1. Specialists in teaching and testing met with teachers of certain subjects to decide upon the educational objectives expected of students.

2. Test exercises were constructed on the basis of the above objectives.

3. Rach test was organized in such a way as to measure the attainment of each objective.

4. The exercises were tried out on groups of college students and were **revised** if the items were ambiguous, lacking in discrimination, or lacking in appropriate difficulty.

5. The revised edition was submitted to group of critics who check the instrument for comprehensiveness in content and in objectives, for accuracy of material, and for validity of exercises.

6. The instrument was revised once again and made available to colleges.

Thus, it can be assumed that the USAFI tests are valid and reliable.

<u>Administration</u>. ^Complete student directions are printed on the cover sheet of the test and preceding each test section. These directions include time limit, explanatory exercises and necessary materials. Specific examiner directions are included in the test manual. These include directions for distribution, time limit, use of materials, and test collection.

The total working time of 120 minutes is not interrupted at any time during the test administration. After the examinee completes one section he goes on to the next section without a stop. The time is adequate.

Material required includes an electrographic pencil for machine-scoring or two No. 2 pencils and an eraser for hand-scoring.

No improvement can be made on the provisions for test administration.

<u>Scorability</u>. Scoring can be done by hand or by machine. A cut-out stencil is provided for hand-scoring. The special answer sheet can be run through the electrographic scoring machine if the special mechanical pencils are used. Handscoring directions are listed on the cut-out stencil.

<u>Interpretation</u>. Percentile and raw score norms are provided. ^The norms for section 1 are based on data from 957 cases in 17 schools; for section 2 on data from 957 cases in 17 schools; for section 3 on data from 890 cases in 17 schools; and for the entire test on data from 956 cases from 17 schools. The test manual provides suggestions for interpretating scores. Colleges are urged to use their own standards.

<u>Conclusions</u>. The test seems to measure the important desired outcomes by means of well-constructed test items. II-A-2 <u>Relevant information</u>. name-EXAMINATION IN BOTANY-COLLEGE LEVEL; author-Examinations Staff for the United States Armed Forces Institute; publisher-American Council on Education; distributor-Cooperative Test Service and Science Research Associates; copyright date-1945; cost-25 test booklets, 2.00, 25 answer sheets, .40, scoring stencils, each, .15, specimen set, .25; forms-form CBo-1-B-4; time-120 minutes; purpose-to measure achievement in the field of college botany.

Format. The test is clearly printed on good heavy quality paper. The six diagrams of the leaf cross-section, ovule, bean fruit and seeds, alga, mushroom, and Marchantia are concisely drawn. There are also two well cut graphs in the test.

<u>Content</u>. 145 5 answer multiple choice items are included in the examination. The test is divided into three sections which consist of the following number of items and which measure the following outcomes:

Section 1 Selection and discrimination of factual

material-70 items

Section 2 Identification and description of structures-

30 items

Section 3 Interpretation of data-45 items

The problems measured in section 3 are comprehensive. They involve the drawing of a conclusion from a list of specific data.

Validity and Reliability. See II-A-1, Validity and Reliability.

Administration. See II-A-1, Administration Scorability. See II-A-1, Scorability

<u>Interpretation</u>. Percentile and raw score norms are provided with the test. Separate norms are listed for sections 1, and 2, and 3. Norms ar also provided for the total test. The data was drawn from 204 cases in 14 colleges. Listed in the suggestions in the test manual is the provision that each college should establish its own local standards.

<u>Conclusions</u>. The section on factual information should be revised to include items with more specific decoys.

CHEMISTRY TEST REVIEWS

II-B-1 Relevant information. name-EXAMINATION IN GENERAL CHEMISTRY-COLLEGE LEVEL; author-Examinations Staff for the United States Armed Forces Institute; publisher-American Council on Education; distributor-Cooperative Test Service and Science Research Associates; copyright date-1944; cost-25 test booklets, 2.00, 25 answer sheets, .40, scoring stencils, each, .15, specimen set, .25; forms-form CCh-2-B-4; time-120 minutes; purpose-to measure achievement at the end of two semesters in general college chemistry.

<u>Format</u>. The test is written on good quality paper. The size of the print is good and the test items are divided by a middle line which aids in the clarity of the general test set up. No diagrams are included. Special answer sheets are required.

<u>Content</u>. The total number of 110 5 answer multiple choice items is divided among three sections. Section 1 contains 35 items which measure general knowledge and information. The section tests definitions of terms, acquaintance with important concepts, and general knowledge of the physical and chemical properties of the common elements and their compounds. Section 2 contains 30 items which measure the application of principles. The section involves the understanding of principles and theories, the ab ility to interpret principles an theories, and the ability to select data involved in making a prediction about the principle or theory. Section 3 contains 20 items which measure skill in balancing equations, chemical stoichiometry, gas laws, molecular weights, and valence. Also included are the quantitative aspects of electrolysis and molarity. Section 4 measures aspects of the scientific method. The distribution of test items is excellent. A laboratory section should be added and diagrams involving problem-solving skills should be included.

Validity and Reliability. See II-A-1, Validity and Reliability.

Administration. Student directions are supplied on the cover sheet of the test. However, these directions are inadequate without the inclusion of specific directions preceding each of the test parts. The time of 120 minutes is not interrupted during the entire examination. The teacher directions listed in the test manual include directions for distribution, time limit, use of materials and test collection.

Explanatory directions preceding each of the test sections should be included for adequate test administration.

<u>Scorability</u>. A cut-out stencil with hand-scoring directions accompanies the test. The test can be machinescored by running the special answer sheet through an electrical scoring machine. Adequate scoring directions are on the stencil and in the test manual.

<u>Interpretation</u>. Percentile and raw score norms are provided for each section of the test and for the total test. These are based on approximately 800 cases which represent 21 colleges. Interpretation directions are indicated in the test manual.

<u>Conclusions</u>. The examination is adequate. ^The addition of a section which measures laboratory skills would add to the value of the examination.

PHYSICS TEST REVIEWS

II-C-1 <u>Relevant information</u>. name - EXAMINATION IN PHYSICS COLLEGE LEVEL; author-Examinations Staff for the United States Armed Forces Institute; publisher-American Council on Education; distributor-Cooperative Test Service and Science Research Associates; copyright date-1945; costsections I and II, 25 test booklets, 1.75, section III, 25 test booklets, 2.00; 25 answer sheets, .40, scoring stencils, each, .15; specimen set, sections I and II, .25 and section III, .25; forms-form CPy-3-B-4; time-sections I and II 90 minutes, section III 90 minutes; purpose-to measure achievement in the field of college physics.

Format. The test is written in good size print on high quality paper. Ther are a total of 25 diagrams which are well cut. Special answer sheets are provided with the test.

<u>Content</u>. The test is divided into three sections. Section I includes 40 items which measure concepts and principles. Section II includes 20 items which measure problem-solving skills. Sections I and II are included in one test booklet. In the second test booklet there are 60 items which make up Section III. This section measures the students ability to think critically. Dubins¹ has divided the test items into groups of subject matter in the following table:

"THE NUM	BER OF Q	UESTIONS	ON THE DIFF.	ERENT
	GROUPS	OF SUBJEC	JT MATTER	
	I	II	III	
Mechanics	7	11	21	
Heat	4	5	11	
Sound	2	3	3	
Light	4	7	8	
Electricity	8	14	17	
totals	25	40	60 *	

1. M.I. Dubins. op. cit. p.66

All sections include 5 answer multiple choice items. The test should include an addition of a section involving laboratory technique.

Validity and Reliability. See II-A-1, Validity and Reliability.

Administration. Student directions are provided on the cover sheet of the examination. However, explanatory exercises are omitted. Clear directions for the teacher are provided in the test manual. The time allotment of 180 minutes seems to be sufficient. The student is directed to work on the test without any interruptions during the working time. Materials include either No. 2 pencils or electrographic pencils. The choice of materials depends on whether the test is to be machine or handscored.

For proper administration, explanatory directions should be provided by the teacher.

<u>Scorability</u>. ^Directions for hand-scoring are listed on the cut-out stencil. For machine scoring the special answer sheet is run through an electrical machine which records by electric impulses the number of correct answers.

<u>Interpretations</u>. No provisions for interpretation accompany the test.

<u>Conclusions</u>. The test can be improved by the addition of a laboratory section and an adequate set of norms.

CHAPTER VI

TRENDS IN TESTING WITH STANDARDIZED INSTRUMENTS

INTEREST IN COOPERATIVE SCIENCE TESTS

This chapter will consider (1) the increase in popularity of the chemistry tests of the Cooperative Test Division, and (2) the present trend in the use of standardized tests as indicated in a survey made by the author.

Annual reports of the College Chemistry Testing Program indicate an increase in the use of standardized tests. King¹ in his 1948 report submitted the following table which summarizes the number of colleges participating in the use of the college chemistry tests:

	"College	Chemistry	Testing	Pro	gram	1938-1948
Year			Number	of	Colle	eges
		Participat	ting		Submi	tting Results
1938		275				149
1939		287				169
1940		289				180
1941	295		x	164		
1942		261				158
1943		112				73
1944		161				46
1945	91			40		
1946	34				29	
1947		121				58
1948		244				216"

The list of colleges submitting results in 1948 included large and small institutions, colleges on the A.C.S. approved

K.C. King. "The 1947-1948 College Chemistry Testing Program ", Journal of Chemical Education, 26:426-428, August, 1949.

list for training chemists, and state and privately supported institutions. The present trend of use of chemistry tests is an upward one as shown by the increase in the popularity of such tests after the war. The recent trend in biology and physics testing has not been indicated in the literature.

PRESENT TREND IN STANDARDIZED TEST USE AS INDICATED

BY AUTHOR'S SURVEY

<u>Procedure used in survey</u>. A questionnaire designed to find out whether standardized or objective-type tests were utilized as course exams or as final exams was sent to the biology, chemistry, and physics departments of 48 major colleges and universities. A random sampling technique was used in the selection of the 48 schools. The following tables based on the results of this survey will indicate the trend in the use of objective-type tests.

Tables based on survey results.

Table 1. RESULTS OF QUERIES ON COLLEGE USE OF OBJECTIVE-TYPE TESTS IN BIOLOGY

COLLE	GES REPLYING	NOT USING	USING
(1)	(2)	(3)	(4)
COURSE EXAMS Number of colleges		6	25
FINAL EXAMS Number of colleges	···· 31	19•4 8 25•8	23 74.2

Table 2. RESULTS OF QUERIES ON COLLEGE USE OF OBJECTIVE-TYPE TESTS IN CHEMISTRY

00	LIEGES REPLYING	NOT USING	USING
(1)	(2)	(3)	(4)
COURSE EXAMS			
Number of colleges		15	19
Percent of colleges FINAL EXAMS		44.1	55.9
Number of colleges		15	19
Percent of colleges		44.1	55.9

Table 3. RESULTS OF QUERIES ON COLLEGE USE OF OBJECTIVE-TYPE TESTS IN PHYSICS

COLLEGES	REPLYING	NOT USING	USING
(1)	(2)	(3)	(4)
COURSE EXAMS	• • 36	8	28
Number of colleges		22.2	77.8
Number of colleges	• 36	13	23
	• 75	36.1	63.9

In all cases the tendency is toward the use of objectivetype tests, both as a measure of achievement during the course and at the end of the course instruction.

CHAPTER VII

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS SUMMARY OF PROCEDURES INVOLVED

<u>Review of the research</u>. The literature was searched for material relative to testing in college biology, chemistry, and physics. All information pertaining to the study was analyzed and digests of the material were made. The digests were organized and presented in sections dealing with college biology, chemistry, and physics tests.

Review of college science tests. The college tests in biology, chemistry, and physics which were constructed since 1938 were obtained from the test publisher. The literature was searched as to the best method of reviewing the tests. The criteria list was chosen as the best means of test reviewing. A suitable set of criteria was formulated and the tests were reviewed according to these criteria.

<u>Present day trend in testing</u>. Questionnaires were sent to departments of biology, chemistry, and physics in 48 colleges and universities. Returns were reviewed for the number of colleges which used exams either to determine achievement during the course instruction or to determine end-of-term achievement. Finally, tables summarizing the returns were constructed.

SUMMARY OF THE FINDINGS

<u>Research in biology testing</u>. Two interesting ideas in test measurement have been intriduced. The first is Fleming's instrument constructed to measure growth in achievement during an orientation course in biology. The second was a test designed by Barnes to measure attitudes about biological phenomena. Other tests which measured progress were those of Mason and Semans. A novel idea was introduced by Hendricks, namely: the development of a circulating library containing test items.

<u>Research in chemistry testing.</u> The development of tests which measure laboratory skills is the new improvement in chemistry testing. Both the paper and pencil test developed by Hendricks and the test involving actual use of laboratory apparatus which was designed by Boeck are good innovations in the area of testing. Other tests were devised, namely: (1) tests which eliminated guessing and (2) tests constructed about historical experiments.

<u>Research in physics testing</u>. Many new type tests were designed in the above atea. Dunning's scientific thinking test which measures the ability to interpret data and to apply principles was well constructed. The stripped-problem test of Schilling was excellent since it dealt with elementary, concrete situations. Weber devised a test similar to that of Boeck's which involved use of laboratory apparatus to determine laboratory skills.

Many others not mentioned contributed to the research in biology, chemistry, and physics testin.

<u>Standardized</u> <u>tests</u>. All recent standardized test have been published either under the direction of the Cooperative Test Division of the Educational Testing Service or under the direction of the United States Armed Forces Institute. Those published by the former are considered to be the best by the author.

CONCIUSIONS

More research has been done on chemistry testing than on either biology or physics testing. No laboratory tests have been discovered in the field of biology. No tests which measure attitudes have been found in the chemistry or physics fields. There is a definite need for the development of tests in biology and physics, for the development of a scientific attitudes test in chemistry and in physics, and for the development of a laboratory test in biology.

The test developed by the Cooperative Test Division are more widely used than those developed by the United States Armed Forces Institute. If more recent editions of the USAFI tests were prepared, they might be more useful.

The use of standardized tests or objective-type tests is on an upward trend as was indicated by the summary on test use by King and by the results of the author's survey.

RECOMMENDATIONS

1. More publishers should enter the field of college science testing.

2. More standardized tests covering all phases of physics and biology should be constructed and published. These new test should include sections which measure laboratory skills and scientific attitudes.

3. Specific objectives to be used in the construction of biology, chemistry, and physics tests should be formulated and published by a committee which represents a good cross-section of the colleges and universities.

4. National norms should be provided for all the college science tests which have been published. Those provided are not adequate for proper interpretation.

5. The USAFI tests should be revised to include more recent information.

6. The Cooperative Test Division is developing a series
of unit tests on various phases in chemistry. Unit tests
should also be prepared for physics and biology courses.
 7. Additional studies dealing with test reviews should
be made in all phases of college education, and high school
education.

BIBLIOGRAPHY

Adams, C.S. "The importance of laboratory work in general chemistry at the college level ", Journal of Chemical Education, XIX (June, 1942), 266-270

American Universities and Colleges. 5th ed., Wash. D.C.: American Council on Education, 1948, pp. 1-1054

Angell, G.W. "Effect of immediate knowledge of quiz results on final examination scores in freshman chemistry ", <u>Journal of Educational Research</u>, XLII (January, 1949), <u>391-394</u>

Ashford, T.A. and Shanner, W.M. "Objective test items of the recognition type that test reasoning and minimize guessing ", Journal of Chemical Education, XIX (February, 1942), 86-89

Barnes, M.W. and Mouser, G.W. " A comparative performance of high school and university freshmen on a test of biological misconceptions ", <u>School</u>, <u>Science</u>, and Mathematics, XLIII (May, 1943), 447-450

Bluh, 0. "Physics exams and the new curriculum ", <u>American Journal of Physics</u>, XVI (January, 1948), 20-24

Boeck, C.H. " A practical examination of skills and techniques acquired in freshman chemistry ", <u>Science</u> Education, XXXI (December, 1947), 320-324

Buros, O.K. The Mental Measurements Yearbook, Highland Park, N.J.: 1941, pp. xxi-674

The Third Mental Measurements Yearbook, New Brunswick, Rutgers University Press, 1949, pp.xiv-1047

Cole, R.D. and von Borgersrode, F. " A scale for rating standardized tests ", <u>School of Education Record</u>, University of North Dakota, XIV (1928), 11-15

Cooperative Test Division of the Educational Testing Service, <u>Cooperative Achievement Tests</u>, N.Y.: Educational Testing Service, 1949, pp. 1-68

Dubins, M.I. " A review of the research in testing in secondary chemistry and physics from 1938-1948 including 14 reviews of standardized tests ", unpublished Master's Thesis, Boston: Boston University School of Education, 1949 Dunning, G.M. "The construction and validation of a test to measure certain aspects of scientific thinking in the area of first year college physics ", <u>Science</u> Education, XXXIII (April, 1949), 221-235

Fleming, M.C. "An evaluation of outcomes of science in higher education ", <u>Science Education</u>, XXVII (November, 1943), 81-85

Foster, L.S. "The 1939-1940 college chemistry testing program", <u>Journal of Chemical Education</u>, XVIII (April, 1941), 159-164

Greene, H.A., Jorgensen, A.N., and Gerberich, J.R. <u>Measurement and Evaluation in the Secondary School</u>, N.Y.: <u>Longmans, Green and Co., 1943</u>, pp. xxvi-639

Haynes, J.H. " An evaluated list of standardized tests in mathematics for grades 9-12 ", unpublished Master's Thesis, Boston: Boston University School of Education, 1948

Hendricks, B.C. "Paper and pencil tests for the laboratory" Journal of Chemical Education, XXII (November, 1945), 543-546

Hendricks, B.C. and Handorf, B.H. "Examination practice in general college chemistry ", Journal of Chemical Education, XV (April, 1938), 176-179

Hendricks, B.C. and Smith, O.M. "Measurable objectives for general college chemistry ", <u>School</u>, <u>Science</u> and <u>Mathematics</u>, XXXVI (October, 1936), 747-752

"Better new examinations from old ", <u>Journal of Chemical Education</u>, XVII (December, 1940), 583-586

Hildreth, G.H. <u>Bibliography of Mental Tests and Rating</u> Scales, N.Y.: The Psychological Corp., 1939, pp.1-295

Supplement, Bibliography of Mental Tests and Rating Scales, N.Y.: The Psychological Corp., 1945, pp. 1-86

Hoff, A.G. <u>Secondary</u> <u>School Science Teaching</u>, Philadelphia: The Blakiston Co., 1947, pp. vii-294

Hurt, W.H. The College Blue Book 5th ed., R.I.: College Blue Book, 1949, pp. 1-464

King, W. " The 1947-1948 college chemistry testing program ", Journal of Chemical Education, XXVI (August, 1949), 426-428 Mason, J.M. " Pre-test and post-test study in biological science ", Journal of Educational Research, XLII (November, 1948), 228-233 Otis, A.S. " Scale for rating tests ", Test Service Bulletin # 13, Yonkers, N.Y.: World Book Co., 1926 Rassweiler. " Improved type of examination for physics courses ", American Journal of Physics, XI (August, 1943), 223-225 Read, J.G. " The Scientific Method ", Unpublished teaching material. Boston University School of Education, 1947 Rinsland, H.D. " A form for briefing and evaluating standardized tests ", Journal of Educational Research, XLII (January, 1949), 371-375 Robertson, G.R. " Examinations in the practice of elementary organic chemistry ", Journal of Chemical Education, XXV (February, 1948), 104-105 Ross, C.C. Measurement in Todays Schools, N.Y.: Prentice Hall Inc., 1947 Rulon, P.J. "Validity of Educational Tests ", Test Service Notebook # 3, N.Y.: World Book Co., 1947 pp. 1-4 Schilling. H.K. " Stripped problems test ", American Journal of Physics, IX (April, 1941), 124-125 Semans, F.M. " The rebuttal test ", School and Society, LIII (March, 1949), 419 Tyler. R.W. " Some findings in the field of college biology ", Science Education, XVIII (October, 1934), 133-142 Weber, L.R. " Dynamic tests for the laboratory ", American Journal of Physics, XII (April, 1944), 101-103 Wellings, R.E. " A new type test in chemistry ", School Science and Mathematics, XXXIX (April, 1939), 351-353 Wright, W.A.E. " The modified true-false item applied to testing in chemistry ", School, Science, and Mathematics, XLIV, (October, 1944), 637-639

APPENDIX

The following is the address of the test publisher whose tests were reviewed:

Cooperative Test Service 15 Amsterdam Avenue New York 23, N.Y.