A STUDY OF COMPUTER-ANALYZED EXAMINATIONS

By

LARRY MAURICE SCROGGINS

Bachelor of Science in Education
Central State University
Edmond, Oklahoma
1966

Master of Teaching
Central State University
Edmond, Oklahoma
1968

Specialist in Education
Pittsburg State University
Pittsburg, Kansas
1971

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF EDUCATION
December, 1977

A STUDY OF COMPUTER-ANALYZED EXAMINATIONS

Thesis Approved:

I. Dwaine Eubanks
Thesis Adviser

Tom E. Moore

John W. Creswell

Thomas D. Johnsten

Norman N. Durham
Dean of the Graduate College

PREFACE

This study is concerned with the development of a method for scaling item analysis data available on general chemistry examination questions to permit prediction of test mean scores when given to general chemistry classes with various grouping characteristics. To accomplish this, three tests were given and a test item file was developed.

The author wishes to express his appreciation to his major adviser, Dr. I. D. Eubanks, for guidance and assistance throughout this study. Appreciation is also expressed to the other committee members, Dr. John Creswell, Dr. Tom Moore, and Dr. Tom Johnsten for their invaluable assistance.

A special note of thanks to my wife, Gail, for her patience and assistance in typing the manuscript. Thanks also is extended to Mrs. Jeanne Mather for helping with the validation of the objectives and to Dr. Bonnie Pride for advice with the statistical procedures.

In addition, appreciation is extended to the instructors of General Chemistry 1314 at Oklahoma State University and to the teaching assistants who administered the test in the General Chemistry 1515 laboratory sections.

TABLE OF CONTENTS

Chapter	Page

LIST OF TABLES

Table		Page

LIST OF FIGURES

Figure | Page

CHAPTER I

INTRODUCTION

The diffusion of computer technology for uses in education has been rapid. A result of this rapid growth has been a lack of structure in defining specific terms used in computer education. Examples can readily be found in which several authors used different terms to describe the same activity. For purposes of this discussion, three areas of computer application will be mentioned. These areas are computer performance in general clerical tasks, computer-assisted instruction, and computer-managed instruction. The area of general clerical tasks includes data storage, examination scoring, and report assembling and will not be discussed further. Other applications are discussed in greater detail in subsequent paragraphs.

Computer-assisted instruction (CAI), a system designed to individualize the educational process, assists the instructor in a remedial/tutorial role, interacting directly with students. The computer can be programmed to give directions, ask questions, and generate examinations in a self-paced instructional format. The computer can provide instructions for students to follow alternative paths to enhance each student's learning. The computer today can be programmed to simulate dialogue. For example, the student may enter information through a typewriter terminal and receive programmed responses via a cathode ray tube or computer-driven typewriter. Simulated laboratory experiences are accomplished by programming the computer to control various learning aids which the student needs at particular times during the learning process. These aids may include tape recordings, film strip projectors, slide projectors, etc. The Federal Government has encouraged the use of CAI systems in educational institutions by providing essential financial support.[1] For convenience in obtaining CAI programs, a periodic publication of such existing programs is published by the Educational Resources Information Center (ERIC), a nationwide information system established by the U. S. Office of Education designed to serve and advance American education.

"When the benefits of CAI are expanded from one subject to a curriculum, from one student to a student body, we begin to approach what can be termed Computer-Managed Instruction (CMI)."[2] CMI provides complete necessary teaching aids and generates test items needed to evaluate student progress. The computer not only individualizes the pace of instruction, but is programmed to control the direction of student learning. CMI also provides a method to improve tests and the instructor's use of tests by: 1) quickly and inexpensively performing statistical calculations, 2) selecting questions according to prespecified criteria, 3) identifying subject areas in which a student is weak, 4) providing learning prescriptions, 5) evaluating course objectives, and 6) evaluating different instructional strategies.

---

[1]Ronald Christopher, "Planting CAI," Educational Technology (August, 1974), p. 59.

[2]Alan B. Salisbury, "Computers and Education: Toward Agreement on Terminology," Educational Technology (September, 1971), p. 38.

Purpose of the Study

The major purpose of this study is to examine the relationship between historical student performance on test file questions and subsequent student performance in general chemistry. Since test files are now routinely shared by various schools and are used by such regional computer centers as the Area Cooperative Computer Educational Systems Services in Des Moines, Iowa, limitations on the transferability of file questions would be useful.

A suitable test file should incorporate questions classified according to the cognitive level if instructional objectives are to be measured. Optimum use of such computer-generated examinations requires that the user understand how the item may be affected when used in different courses. The maximum value of file questions may depend on whether individual questions can be generalized to the achievement of instructional objectives in different courses. This study does not attempt to generalize beyond a single discipline nor to generalize within a discipline; only general chemistry classes are considered. The instructors, textbooks, instructional strategies, student characteristics and course objectives may, however, be different. These characteristics may be compensated for by prediction from a linear relationship between the test question file and different types of general chemistry courses.

A minor purpose of this study is to determine if different instructors can consistently use Bloom's <u>Taxonomy of Educational Objectives</u>[3] to place questions into the same classifications. The relative degree of

---

[3]Benjamin S. Bloom, et al., <u>Taxonomy of Educational Objectives</u> (New York, 1956).

mastery of individual test item scores will indicate the degree to which the various cognitive skills have been developed. The same examination can be used to monitor the level of difficulty of the test which is represented by the test mean score and measure the achievement of instructional objectives; thus, the test is serving a dual purpose. One purpose of the achievement test is to monitor the degree of student achievement. The second is to measure the degree of development of the cognitive skills.

Statement of the Problem

Subject to the limitations stated later, this study has two objectives. One objective is to determine the limitations of the transferability of carefully prepared test files. This should allow test files to be shared by different educational institutions and also allow instructors to feel relatively sure that the same level of difficulty is being assessed. A second objective of this study is to determine if Bloom's taxonomy can be used by different instructors to place chemistry questions into the same classifications. This should provide an instructor with a method for classifying chemistry questions for the purpose of test construction regardless of the area or locality in which the questions will be used. The test file questions can be used to assess achievement of instructional objectives in general chemistry classes in which the students have different educational backgrounds.

Value of the Study

This study will: 1) provide the instructor with criteria to produce examinations with different items, the items each having the same relative level of difficulty, and assessing the same types of learning, 2) allow the instructor to retest those learning skills a student fails to master without repeating entire tests, 3) permit the instructor to use the same question file to produce tests of predetermined difficulty for various chemistry classes grouped according to different characteristics such as major field of study, 4) help the instructor assess the learning skills being mastered by his students and, thus, allow him to revise his teaching strategies as necessary and provide prescriptions for remedial work, 5) allow the student to retain old tests to be studied without reducing test security, and 6) eliminate untried and possibly poor test items from examinations.

Assumptions

The control group involved in this study is assumed to be a random sample of undergraduates, all being classified as science or engineering majors. It is assumed that student performance on test items will approximate a Gaussian distribution.

It is assumed that different course instructors will not cause significant differences in the mean scores on any individual test item from one semester to the next, so long as course objectives and student selection criteria remain the same.

It is assumed that computer storage of item characteristics may permit random question selections to produce examinations having predetermined characteristics if the student group being examined has met the same selection criteria as the original student group.

It is assumed that selection criteria (educational background, major field of study, etc.) used for placing students in Chemistry 1314 and Chemistry 1515 have remained the same, otherwise, variations would not be significant.

Limitations

In the general chemistry course for engineering or science majors, student attitudes toward taking the test were influenced by the knowledge that the examination was conducted as part of a study rather than as part of course grades. Thus, student motivational level was decreased exhibiting a Hawthorne effect. According to Kerlinger in his book, Foundations of Behavioral Research, "Almost any change, any extra attention, any experimental manipulation, or even the absence of manipulation but the knowledge that a study is being done, is enough to cause subjects to change."[4]

The results of this study are limited to the population of students used. The study utilized students enrolled at Oklahoma State University in the first general chemistry course for science or engineering majors (1314) and three sections of the second general chemistry course for engineering or physical science majors (1515). A small student population of limited background enrolled at the University of Science and Arts of Oklahoma in general chemistry was used for the pilot study. Another limitation of the study is that it does not address the effects of instructor and textbook variables.

The final limitation to this study is that tests 4A and 4B must not be the same. If tests 4A and 4B were the same, they would have no predictive value.

---

[4] Fred N. Kerlinger, Foundations of Behavioral Research (New York, 1964), p. 318.

Definition of Terms

Chemistry 1124

A chemistry course at the University of Science and Arts of Oklahoma in which students are not grouped according to declared majors. In the pilot study, by accident, rather than by design, the students had all declared themselves to be science majors.

Chemistry 1314

The first general chemistry course at Oklahoma State University for science or engineering majors.

Chemistry 1515

The second general chemistry course at Oklahoma State University for students who have declared themselves to be science or engineering majors.

Coefficient Correlation

This is a coefficient between 1 and -1 that measures the degree of fit between sets of data to a least-squares single variable linear regression line.

Coefficient of Determination

The square of the simple correlation coefficient. It measures the closeness of fit of two variables to a regression line and has a value between zero and one. Coefficient of determination values approaching one characterize a good fit between the two variables.

Control Group

Those subgroups of students of general chemistry that have declared a science or engineering major.

Difficulty Index

The fraction of students responding correctly to a test item or group of items.

Discrimination Index

The discrimination index for an item has the value of (t-b)/n, where t is the number of students in the top 27% of the class (as determined by total score) who responded correctly, b is the number of students in the bottom 27% of the class (as determined by total score) who responded correctly, and n is the number of students in 27% of the class.

F-test

The standard deviation of each set of data is squared. The standard deviation that has the larger numerical value is divided by the standard deviation of the lowest numerical value to give an f-ratio value.

Ideal Test

A test designed from item analysis and objectives to assess both the level of difficulty and the degree of mastery of the objectives.

Item Analysis

The generic term given to statistical analyses of item characteristics.

Item Mean Score

The average of all the test item scores.

Pilot Study

The preliminary study conducted at the University of Science and Arts of Oklahoma to determine the appropriateness of the methods to be used in the principal study. The pilot study included the same prediction methods as the principal study, but a very small population of students was used.

Split-half Forms

Two forms of an examination containing identical test items where the items are reversed in meaning.

T-test

Deviation of the estimated mean from the actual mean equal to the standard deviation divided by the square root of the sample size of the unit.

Taxonomical Objective

Classification of cognitive level required of student to exhibit mastery of subject matter.

Test Mean Score

The average of all the item scores.

Types of Learning

The objectives which deal with the recall or recognition of knowledge and the development of intellectual skills and abilities.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

During the 1960's as enrollments in institutions of higher education rapidly increased, an effort was made to modify instructional strategies to respond to the individual needs of students while meeting the economic necessity of mass instruction. As large amounts of money were provided to support educational innovation during this period, more and more use of the computer was made to enhance the quality and availability of education and to transfer the responsibility of learning from the instructor to the student.

The idea of self-instruction began as early as 1954 when Dr. B. F. Skinner of Harvard invented a mechanical machine whereby questions were presented on tape viewed through a window on the machine. A response was given by the student who then received immediate feedback. In the latter 1950's an experiment was conducted using an IBM 650 computer which was connected to a typewriter terminal. Arithmetic problems were stored in the computer and typed out and the students typed in the answers. "The way was paved for self instruction via the computer. Likewise, the seed of an idea-time sharing had been planted."[1]

---

[1]Justine Baker, The Computer in the School (Bloomington, 1975), p. 18.

The computer, however, did not become an immediately effective tool in education because of the complexities of programming languages. It was not until John Backus and Irving Ziller developed FORTRAN (an automatic programming language) that the computer was further utilized as a means of reorganizing and administering knowledge.

Educational Systems Using Computers

In 1960 the first major computer-assisted instructional system was created under the supervision of Dr. Donald Bitzer at the University of Illinois. The project was called PLATO (Programmed Logic for Automatic Teaching Operation). In 1961 the first teaching attempt using computerized programmed instruction in higher education was made by PLATO, and by 1965 the first complete college course was administered, solely by the PLATO system. The time-sharing method of computer-assisted instruction was added to PLATO in 1968.

The PLATO system evolved through four developmental phases from PLATO I in 1960 to the PLATO IV system now being used. The system is controlled by a large computer with 450 terminals. Each PLATO IV terminal consists of a keyset for communication between the user and the computer, and a display panel for showing graphic information and color slides.

The National Science Foundation has played a supportive part in furthering the use of computers in education. Computer networks, established for the purpose of sharing computer resources for instructional purposes, have been implemented by the National Science Foundation. By 1975 some 30 regional computing networks had been established in 300 institutions of higher education and some secondary schools. Another project, Computer Home Delivery Systems, supported by the National Science Foundation through the MITRE Corporation was installed in the town of Reston, Virginia to test the feasibility of computer-controlled television for use within the home. This system was designed for use by the handicapped, elderly, and those with special learning problems.

In 1970 TICCIT (Time-Shared Interactive Computer-Controlled Information Television) was initiated at the Institute for Computer Uses in Education at Brigham Young University. The philosophy of the TICCIT project was "to plant academically successful and economically feasible CAI systems in American schools and to change the role of the classroom teacher to that of tutor-advisor, diagnostician, and problem solver."[2] It was perceived that since community colleges were relatively new in American education and should be receptive to innovation, this would be a logical place to test this system. Programs in mathematics and English designed specifically for the community college level were implemented at Phoenix College and the Alexandria campus of Northern Virginia Community College.

In an article written by Cynthia L. Heinje, an instructor at Rhode Island Junior College, Warwick, another example is cited of a junior college paving the way in providing instructional support services to education. The services are available not only at the College but also to people throughout the State of Rhode Island. Through the computer-assisted instructional approach programs, many subjects are presented through computer terminals, slides, sound tapes and written material. The programs instruct and test students. Immediate feedback is avail-

---

[2]Ibid, p. 22.

able providing for correction of errors. The student may progress at a self-imposed rate of speed. Each student works individually from a typewriter terminal, and an instructor gives assistance as needed. There are over 30 terminals through the college. A central laboratory area is provided where students may select programs on different subjects. "During the 1973-74 academic year, computer terminals provided over 5,000 hours of instruction."[3]

Under the term "Computer-Managed Instruction" at Rhode Island Junior College, pretests are given in a specific subject area and then the students are directed from there, perhaps to a film, tape demonstration, etc. After finishing the assignment the student returns to the terminal and takes a post-test over the material to determine proficiency. An evaluation progress is built into the system whereby instructors can monitor the progress of students in all subject areas.

Another facet of the system at Rhode Island Junior College is the computerized Career Information Service.

> The basic purpose of this system is to provide college students, high school students and others in Rhode Island with brief, immediate and easily accessible information on careers so that valuable time can be saved and unnecessary research eliminated.[4]

Terminals are located throughout the state and are linked via telephone lines. Each terminal consists of a cathode-ray tube on which information is displayed, a keyboard for entering student responses and a copier to duplicate information displayed on the screen. This gives information about occupations, college life at 1600 colleges, appren-

---

[3]Cynthia L. Heinje, "Junior College Computerized Instructional Support Services," Educational Technology (August, 1975), p. 32.

[4]Ibid., p. 33.

ticeship programs in skilled trades, military life, etc. One of the most important segments of the system is a listing of immediate job openings in Rhode Island.

A recently developed phase of computer instruction, the Teaching Information Processing System (TIPS), has proven to be very effective in individualizing instruction in the large classroom, where individual evaluation of student progress is difficult. Since large classroom instruction must cater to the average student, leaving the bright students unchallenged and those with learning difficulties undetected, TIPS provides both the student and the instructor with an opportunity to determine each student's proficiency before formal examinations. The professor prepares and administers surveys throughout the course in the form of short, multiple-choice or objective-type questions. As a result of these surveys, individual reports are prepared for each student. TIPS contradicts the feeling of some educators that computers depersonalize education because this system permits the professor to treat each learning problem on an individual basis.

"TIPS is most appropriate to disciplines where the subject matter objectives are reliably measured by well-formulated, objective-type questions...and...in those areas where the subject matter is cumulative in nature."[5] Feedback on TIPS research programs has shown that the high achieving students increase their performance by approximately 13% while low achievers increase performance by approximately 19%.

The computer has proven to be an indispensable tool in the many functions involved in the educational process. Many authors use dif-

---

[5]"TIPS, a diagnostic tool to individualize instruction in the large class," p. 6.

ferent terms to describe these activities. Alan B. Salisbury in his article "Computers and Education: Toward Agreement on Terminology" grouped the applications of the educational uses of computers into three areas: "Administrative, Ancillary and Instructional."[6]

Administrative computer functions are "those performed in direct support of the administrator element."[7] These may include but not be limited to payroll, record keeping, class scheduling, student rosters, financial reports, library management, guidance programs in vocational and curriculum planning, and current problems such as drug abuse, etc. Some data banks and information retrieval systems could fit into this category.

Ancillary computer functions are those functions which "serve equally well the learner, author-teacher and administrator elements of the instructional system."[8] This function consists primarily of problem solving, the oldest educational computer use. Ancillary functions may be scientific or business in nature; however, since they are used by educators, they are considered educational functions.

Instructional computer functions include "all applications in which a computer is used in direct support of an instructional function involving subject matter, instructor and student(s)."[9] The most widely used term for the instructional functions is "computer-assisted instruction" (CAI). Some other terms used by some authors include computer adminis-

---

[6]Alan B. Salisbury, "Computers and Education: Toward Agreement on Terminology," <u>Educational Technology</u> (September, 1971), p. 35.

[7]Ibid., p. 36.

[8]Ibid.

[9]Ibid., p. 39.

tered instruction, computer-aided instruction, computer-simulated instruction, and computer-managed instruction.

Test Design Using Computers

Lippey suggests in an article in Educational Technology, "In our enthusiasm to apply technology to aid in the instructional process, we may have overlooked some simple and direct functions that can provide valuable service at very low cost."[10] By utilizing the computer, the instructor may be provided with tests designed to meet the various needs of the students.

Paul Ansfield, as early as 1969, wrote a program for the University of Wisconsin which would allow items to be selected according to the following characteristics:

1. Item subject matter catalog.
2. Item number.
3. Subject matter.
4. Correct answer.
5. Type of question.
6. Item analysis information.
7. Difficulty level.
8. Discrimination ability.
9. Random selection.[11]

The real benefits of the computer generated examinations, according to Ansfield, are their neatness and legibility, together with zero error probability. Many different examinations can be provided, freeing the instructor's time and maximizing examination security.

The key characteristic from this set is different examinations.

---

[10] Gerald Lippey, "The Computer Can Support Test Construction in a Variety of Ways," Educational Technology (March, 1973), p. 9.

[11] Paul Ansfield, "A User Oriented Computing Procedure for Compiling and Generating Examinations," Educational Technology (March, 1973).

The examinations must be equivalent but still retain their individuality. Thus, one must control two factors, the degree of achievement and the type of learning, and still use different questions on each.

Ebel reports that "if modern knowledge and techniques of test construction are applied, most educational achievement tests can be made to yield scores having reliability coefficients that at least approach 0.90."[12] It appears from this that a test could be constructed to give rather good results whose reliability coefficients should be acceptable for most decisions. According to Mehrens and Lohmann:

> Although there is no universal agreement, it is generally accepted that tests used to assist in making decisions about individuals should have reliability coefficients of at least .85. For group decisions, a reliability coefficient of about .65 may suffice.[13]

The item analyses are available but have been, to a large extent, neglected by educators. The item analyses may be chosen in such a way as to eliminate poor questions and still provide a relative degree of scoring reliability. The item analyses provides data that can be used to construct tests and produce desired results. The items on the test are recommended to have an average difficulty of 62.5% and vary in difficulty from 15 to 85%. Sax and Cromack also suggest that "if time limits for the test are generous, as they usually should be for achievement tests, the order of presentation of the items has little effect on

---

[12] Robert L. Ebel, Essentials of Educational Measurement (Englewood Cliffs, 1972), p. 408.

[13] William Mehrens and Irvin Lohmann, Standardized Tests in Education (New York, 1969), p. 41.

student scores."[14]

Criterion-Referenced Measures

Various attempts have been made to validate the use of Bloom's Taxonomy of Educational Objectives.[15] In a 1966 study, Russell Kropp conducted a study "to test empirically the structure of the handbook Taxonomy of Educational Objectives 1 -- Cognitive Domain."[16] The study generally supported the hypothesis that there exists an inverse relationship of mean performance and taxonomic level. However, for the science forms there was a systematic reversal of means on the synthesis and evaluation subtests. Thus, as the mean taxonomical level of objectives of questions comprising a test increases, the mean score on the test decreases. Kropp could see a serious problem if taxonomy-type tests were constructed and come into wide use. Kropp stated that the problem, establishing norms for taxonomy-type tests, was quite complex and probably cannot be solved by current norming techniques.

Robert Geisinger used the radex method for determining the construct validation of the Bloom's taxonomy. The researchers reasoned that if Bloom's taxonomy is indeed hierarchical in nature, then radex structure should be observed according to theory. The radex was plotted by a computer program as concentric circles that increased toward the outer

---

[14]Gilbert Sax and Theodore R. Cromack, "The Effects of Various Forms of Item Arrangement on Test Performance," Journal of Educational Measurement, Vol. 3 (1966), p. 309.

[15]Bloom.

[16]Russell P. Kropp, "The Construction and Validation of Tests of the Cognitive Processes as Described in the Taxonomy of Educational Objectives," (September, 1966).

circle in order of higher complexity. "Knowledge variables were somewhat anomalous but evidence for radex structure was reported in graphs of the data obtained."[17]

Criterion-referenced examinations can be used to assess the effectiveness of teaching strategies. Friedman in 1977 conducted a study in which group A emphasized certain cognitive levels and group B deemphasized those same cognitive skills. "Achievement scores at the memory and application levels of pupils taught by the respective A-group teachers were significantly higher than scores of the corresponding B-group pupils."[18] According to this study a significant positive difference exists between teacher emphasis of the memory and application levels and pupil achievement at those levels.

The classification of test items makes them more useful than just a determiner of student performance. The renewed interest in criterion-referenced measures is, in part, a response to the evaluation of mastery learning and individualized instruction. Payne, in 1974, listed the following eight general uses for criterion-referenced measures:

> "Placement in a Learning Continuum, Diagnosis of Individual Student Achievement, Monitoring of Individual Student Progress, Diagnosis of Class Achievement, Monitoring Class Progress, Evaluation of Curricula, Project and Program Evaluation, Grouping on the Basis of Content Achievement."[19]

---

[17]Robert W. Geisinger, "Construct Validation of Hierarchical Taxonomies of Educational Objectives," Educational Technology (May, 1973), p. 28.

[18]Morton Friedman, "Teachers' Cognitive Emphasis and Pupil Achievement," Educational Research Quarterly (Spring, 1977), p. 42.

[19]David A. Payne, The Assessment of Learning, Cognitive and Affective, (Lexington, 1974), p. 302.

Criterion-referenced measures are intended to measure what, not how much, the student has or has not learned. They can provide information not readily available from other sources and can be used to supplement but not supplant existing methodology.

The trend to develop area wide test files is still running rapid. The increasing availability of banks or repositories of performance objectives and test items, and the development by testing companies in this area, have increased the likelihood that criterion-referenced tests will be available and used for measuring school achievement. Tuckman reported that "such criterion-referenced tests, as they appear, will have the advantage of allowing each school district to target its testing program to its own goals and to monitor goal attainment in an absolute rather than a relative sense."[20] In 1974 Kenneth Fast recommended that test items be assembled from several subject areas to provide a test bank from which classroom examinations could be constructed.[21] The items were to be selected from the first four cognitive levels and stored in a computer for easy access. Classification of items in the affective domain should be neglected.

Educational Objectives Using Computers

If the examination scores can be predetermined on another examination from analysis of those items included, it would be impossible to

---

[20]Bruce W. Tuckman, Measuring Educational Outcomes (New York, 1975), p. 296.

[21]Kenneth V. Fast, "An Analysis and Classification of the ACS-NSTA High School Chemistry Achievement Tests Using Bloom's Taxonomy-Cognitive Domain," Science Education Vol. 58 (1974).

determine if the different tests measured the same kind of learning. A student may be able to answer a question on terminology, a definition or theory but not be able to answer a question where application is involved.

Test questions can be written to measure the level of mastery of the material as well as the development of cognitive skills.

Bloom has attempted to classify all the educational objectives which deal with the recall or recognition of knowledge and the development of intellectual skills and abilities associated with learning in the cognitive domain. In the February, 1975 issue of Educational Technology, Schonberger asserts that test item pools with easy-to-produce questions on terminology, etc. will assure mediocrity and turn off brighter students.[22]

A good test should contain questions based on two criteria. The questions need to be compiled on a level of difficulty so that slow, average, and bright students' knowledge can be tested. The questions need to test each student's development of cognitive skills. Reliable categorization of the level of the mental process required to correctly answer any given question is difficult because variations in the manner of instruction may affect the student's ability to respond correctly.

In an article "Course Improvement Through Evaluation," Lee J. Cronbach stated:

> The distinction between factual tests and tests of higher mental processes, as elaborated for example in the Taxonomy

---

[22]Richard J. Schonberger, "Modular Instruction with Computer-Assembled Repeatable Exams: Second Generation," Educational Technology, (February, 1975), pp. 36-38.

> of *Educational Objectives*, is of some value in planning tests, although classifying items as measures of knowledge, application, original problem solving, etc., is difficult and often impossible. Whether a given response represents rote, recall, or reasoning depends upon how the pupil has been taught, not solely on the questions asked.[23]

There are many variables in dealing with achievement test questions. A categorization of these questions by the thought process is involved in answering the question. Any investigation should take into account the various variables in attempting to explore the effects of any one or several variables in a test situation. Thus, before a question involving comprehension or application could be answered, various types of information must be available.

Item analysis provides a good method for selection of good questions of appropriate difficulty. A reasonable method for selecting questions to measure the types of learning involved is to classify selected questions according to Bloom's *Taxonomy of Educational Objectives*,[24] as set forth in Table I.

This would enable the examiner to incorporate both properties into the test, thus, giving the test validity and reproducibility.

Prediction of Achievement

A review of recent measurement literature (1967-1976) prior to this study indicated that no study had been reported in the literature on test files being used to establish the relationship between student performance on test file questions and subsequent student performance in gen-

---

[23] Lee J. Cronbach, "Course Improvement Through Evaluation," *Teachers College Record*, Vol. 62 (1963), p. 672.

[24] Bloom.

TABLE I

BLOOM'S TAXONOMICAL OBJECTIVES

Knowledge

1.00 Knowledge
- 1.10 Knowledge of Specifics
  - 1.11 Knowledge of Terminology
  - 1.12 Knowledge of Specific Facts
- 1.20 Knowledge of Ways and Means of Dealing with Specifics
  - 1.21 Knowledge of Conventions
  - 1.22 Knowledge of Trends and Sequences
  - 1.23 Knowledge of Classifications and Categories
  - 1.24 Knowledge of Criteria
  - 1.25 Knowledge of Methodology
- 1.30 Knowledge of the Universals and Abstractions in a Field
  - 1.31 Knowledge of Principles and Generalizations
  - 1.32 Knowledge of Theories and Structures

Intellectual Abilities and Skills

2.00 Comprehension
- 2.10 Translation
- 2.20 Interpretation
- 2.30 Extrapolation

3.00 Application

4.00 Analysis
- 4.10 Analysis of Elements
- 4.20 Analyses of Relationships
- 4.30 Analysis of Organizational Principles

5.00 Synthesis
- 5.10 Production of a Unique Communication
- 5.20 Production of a Plan, or Proposed Set of Operations
- 5.30 Derivation of a Set of Abstract Relations

6.00 Evaluation
- 6.10 Judgments in Terms of Internal Evidence
- 6.20 Judgments in Terms of External Criteria

---

eral chemistry. Wasik reported an attempt to predict student success in statistics from student knowledge of algebra. He reported that "algebra is strongly related to performance on statistics and that a positive self-concept can also be a determinant of success in statistics."[25] This suggests that prediction can occur across interdisciplinary lines.

Standardized tests have been used to predict grades. Michael, Knapp and Young reported a study for the purpose of determining the validity of TABS (Tests of Achievement in Basic Skills) in predicting course grades. Correlation of the total TABS score against teachers' marks were .63, .73 and .71. "These correlations were considered to be at a level consistent with, and in many cases superior to, results obtained in previous studies relating test scores to achievement as reflected in teachers' marks."[26] Two attempts were found to predict student performance from different kinds of achievement tests.

French and Ryan's 1976 study reported that "generally with the possible exception of the low SES (socio-economic level) schools, the achievement test was as valid a predictor as was the verbal intelligence measure, and both of these measures were as effective for short-term predictions as were teacher grades."[27] An investigator certainly should

---

[25]John L. Wasik, "Prediction of Success in a Behavioral Science Course," Improving College and University Teaching

[26]James Young, Robert Knapp, William Michael, "The Validity of the Tests of Achievement in Basic Skills for Predicting Achievement in General Mathematics and Algebra," Educational and Psychological Measurement, Vol. 30 (1970), p. 953.

[27]James J. Ryan and James R. French, "Long-Term Grade Predictions for Intelligence and Achievement Tests in Schools of Differing Socio-Economic Levels," Educational and Psychological Measurement, Vol. 36 (1976), p. 558.

be aware of the effect of SES on the prediction of student performance and consider this when establishing the design of an investigation.

Goolsby reported a study on the success of CLEP (College Level Examination Program) to predict student performance as represented by their grade point average at the sophomore and junior years in college. The correlation between student's grades in physical sciences and the CLEP test score on the CLEP physical science subtest was very low. "Coefficients of correlation between CLEP and certain subject areas are quite low for any practical predictive use."[28] According to Goolsby "a rigorous determination and definition of curricular objectives and the construction of criterion measures (cognitive and affective) are necessary for higher education to meet its responsibilities for selection, placement, and advisement."[29] Locally developed and administered examinations may be the best predictor of student performance.

Mueller and Loeb studied the use of a scale of high schools in predicting college grades. The method used was to scale high school grades on the basis of grades earned in college. This was accomplished through linear regression and "this scaling method, for 13 colleges, netted an increase of .11 in the median correlation of high school and college grade averages."[30]

Classical linear-prediction may be a good method to use to predict

---

[28]Thomas M. Goolsby, Jr., "The Validity of the College Level Examinations Programs Tests for Use at the College Sophomore Level," _Educational and Psychological Measurement_, Vol. 30 (1970), p. 376.

[29]Ibid.

[30]Jane W. Loeb and Daniel J. Mueller, "The Use of a Scale of High Schools in Predicting College Grades," _Educational and Psychological_ Measurement, Vol. 30 (1970), p. 381.

student performance. Kleinke used linear-prediction to determine student performance. He used a 100-item five option-multiple choice test to determine student performance on a larger test. The questions were chosen from the larger test by matrix-sampling because the sampling was across examinees as well as across items. The linear-prediction was highly accurate. There was no discrepancy as great as four percentile points. "It should here be noted that although linear prediction proved effective with these data, it is not claimed that the generalizability of the technique has been established. Further work, both empirical and theoretical, is indicated."[31] This research study adds new evidence to support or reject the linear-prediction method.

Summary

During the last 20 years the computer use for compiling, analyzing, and transferring knowledge has progressed. By utilizing the computer, an instructor is able to provide more individual assistance to those who have learning difficulties and special interests. Further, more meaningful and applicable means of determining student proficiency and of providing examination security are also available through the utilization of the computer.

---

[31] David J. Kleinke, "A Linear-Prediction Approach to Developing Test Norms Based on Matrix-Sampling," Educational and Psychological Measurement, Vol. 32 (1972), p. 75.

CHAPTER III

EXPERIMENTAL DESIGN

Introduction

The three major topics treated in this chapter are the development of the test file, the pilot study, and the principal study. The test file consisted of 500 questions chosen from examinations previously administered to students at Oklahoma State University. Each question was chosen on the basis of four criteria: Each question possessed: 1) a discrimination index of +.25 or greater, 2) a difficulty level between 25% and 90%, 3) three distracters, all of which were operating, and 4) the multiple-choice type format so that it could be easily computer scored. To lend flexibility to the test file and broaden its use, the questions were classified according to Bloom's taxonomical objectives, thus, not only the degree of mastery of the material as reflected by the mean test scores could be ascertained but also the degree of mastery of various cognitive skills could be determined.

The primary purpose of conducting the pilot study was to use a small population of students to run a tentative check on the major theme of the principal study. The major theme was that test files are transferable from one chemistry class to another and from one university to another. The pilot study consisted of a 30-question multiple-choice examination administered to students at the University of Science and

Arts of Oklahoma. The first ten questions were used to establish the linear prediction line and they were correlated with the second ten questions.

The purpose of administering the first two examinations of the principal study was to determine the reliability of a large number of questions selected from the developed test file. These examinations were hour-long examinations consisting of multiple-choice questions conducted at Oklahoma State University in general chemistry for science or engineering majors. The reliability of the test file questions was established by comparing test mean scores on split-half forms and by comparing test file item scores and actual test item scores. Before any reasonable attempt could be made to determine the transferability of the test file questions to either different groupings within the same academic discipline or to equivalent classes in other universities, the reliability of the individual item scores and test mean scores had to be determined. Questions with low reliability would not be good predictors.

The principal study was continued with the administering of a third examination to determine if test scores on individual items from the test file developed for science or engineering majors could be scaled using linear regression prediction to produce the same mean test scores when given in General Chemistry 1515 for physical science or engineering majors. This examination was an hour-long examination composed of multiple-choice type questions and was given to students majoring in physical science or engineering at Oklahoma State University.

Development of the Test File

Multiple-choice type questions from previously given examinations at Oklahoma State University were scrutinized and those questions deemed to have appropriate discrimination indexes and difficulty levels were classified according to taxonomical objective and incorporated into a comprehensive test file. For purposes of organization and convenience the questions were grouped according to subject headings chosen to correspond as closely as possible to the subject areas covered in the spring, 1975, semester at Oklahoma State University and numbered sequentially (Table II). There were two reasons for using this particular grouping method. One reason was that since the principal study was to be conducted at Oklahoma State University, it would be more convenient to have the major subject categories parallel with those presently being taught at Oklahoma State University. Since the order of presentation of the theory or subject groupings varies from one class to another and from one semester to another, the investigator's second purpose was to group the questions in such a manner that questions could be readily extracted from the file for future use.

TABLE II

CATEGORIZATION OF TEST FILE QUESTIONS

Chemistry 1364

1. Introduction to the Study of Chemistry
2. Measurements
3. Matter
4. Structure of the Atom
5. Structure of Compounds
6. Periodic Classification of Elements
7. Chemical Nomenclature

Table II (Continued)

8.	Calculations of Formula or Molecular Masses (Stoichiometry)
9.	Gases
10.	Chemical Equations
11.	Calculations Involving Chemical Equations
12.	Water
13.	Solutions and Colloids
14.	Acid, Bases, Ionic Equations
Chemistry 1474	
1.	Mixtures
2.	Thermodynamics
3.	Kinetics
4.	Chemical Equilibrium
5.	Acids and Bases
6.	Oxidation-Reduction
7.	Organic Structure
8.	Chemistry and Environment

Distracters

The questions included in the developed test file were of the multiple-choice format and contained three working distracters. Hedges,[1] in his article, "How to Construct a Good Multiple-Choice Test" suggested the use of four distracters and argued that the reduction of the number of options to three or four increases the element of chance and reduces the reliability of the test. However, since the questions to be included in the test file were questions previously constructed for use at Oklahoma State University, the investigator had no choice in determining the number of distracters. In selecting questions for inclusion

[1]William D. Hedges, "How to Construct a Good Multiple-Choice Test," The Clearing House (September, 1964), p. 10.

in the test file, the investigator followed the other criteria in determining the characteristics of good multiple-choice questions as suggested by Hedges in the following ten rules:

Rule 1:	All options should be grammatically consistent.
Rule 2:	Make the stem long and the distracters or options brief.
Rule 3:	All extraneous material should be excluded from the stem.
Rule 4:	The stem of each test question should contain a central problem.
Rule 5:	Double negatives should be avoided.
Rule 6:	The optimal number of options is five.
Rule 7:	All distracters should be plausible.
Rule 8:	The more homogeneous the options, the higher the level of understanding required.
Rule 9:	The correct response should not be consistently longer or shorter than the decoys.
Rule 10:	Only one of the options should be the correct or best answer.

Discrimination Indexes

Each test in this study was designed to produce test mean scores within $\mp$ .5% of the theoretical 50% maximum discrimination level. The discrimination indexes of the questions varied between +.25 and +.90. Most educators agree that any question whose discrimination index is below +.25 should be rejected, therefore, +.25 was chosen as the lower limit for the inclusion of questions in the test file. The upper limit of +.90 for the discrimination index was not chosen but was simply the highest observed discrimination index among the questions used.

Difficulty Levels

Test items having difficulty levels between 25% and 90% were selected for inclusion in the test file. Arthur Storey, in his article "The Measurement of Classroom Learning" listed several strong arguments for rejecting questions below the 25% difficulty level. He wrote:

> Items at or below the 25% level of difficulty are discarded because they are (1) too difficult for the group concerned; (2) likely to be unduly influenced by guessing; (3) will likely fail to discriminate since good students are generally no more skilled at guessing than poor ones; and (4) even if the very difficult items should discriminate satisfactorily, they do so on the basis of too few examinees.[2]

Dorothy Wood suggested in her book Test Construction that the level of difficulty be varied or controlled in the range between +15% and +85%.[3] The reason for varying the level of difficulty of the questions was so that a certain percentage of easy questions could be included for the lower scoring students and a certain percentage of more difficult questions could be included for the higher scoring students. A test constructed to produce a mean score of 50% gives maximum discrimination. The mean score is a measure of the overall difficulty level of the examination.

Design of the Pilot Study

The central purpose for conducting the pilot study was to use a very small population to test the major theme of the principal study that test files are transferable. If the results supported the hypothesis, this would justify the principal study. The basic idea proposed by the pilot study was that the test file questions taken from Oklahoma State University could be scaled using linear regression and produce predictable scores when used at the University of Science and Arts of Oklahoma. The pilot study incorporated the use of an examination consisting

---

[2] Arthur Storey, The Measurement of Classroom Learning (1970), p. 87.

[3] Dorothy Wood, Test Construction (Columbus, 1960).

of 30 multiple-choice questions (Appendix A). The questions were constructed in such a way that the discrimination indexes, difficulty levels, and taxonomical objectives were controlled. The examination was sub-divided into three groups of ten questions each. The groups were referred to as Tests 1A, 1B and 1C. The mean indexes of discrimination for each group were +.38, +.44 and +.44, respectively. The levels of difficulty for each group were 66.8, 66.8 and 66.6%, respectively.

The examination was also designed to measure the mastery of cognitive skills. In an article entitled "Constructing Tests with the Mentrey Tutorial Testing System," Frieda Libaw[4] gave a fifty item examination using objectives. The relative distribution of these objectives are given in Table III. The real significance Libaw contributes to this study resides in the fractional percentages allowed for each objective (learning, comprehension, and remaining cognitive skills.) Thus, two questions from the knowledge category, two questions from the comprehension category, and six questions from the application category were incorporated into each set of ten questions (Table IV).

TABLE III

EXAMPLE USE OF 50 ITEMS

25% rote learning
25% comprehension
50% remaining cognitive skills

TABLE IV

EXAMPLE USE OF 30 ITEM EXAMINATION DIVIDED INTO SETS OF TEN

	No. of Questions for each set of Ten
20% Knowledge	2
20% Comprehension	2
60% Application	6

Construction of the Principal Study

Reproducibility of Difficulty Levels

The purpose of this phase of the study was to determine if reproducible results could be obtained from split-half test forms having identical questions arranged in different order. This would establish the reliability of the questions in the test file. The difficulty levels must be stable from one semester to the next to be useful. This was accomplished by comparing the test scores on individual questions from the test file with the observed performance scores on the split-half test forms.

The instructors of general chemistry courses for engineering or science majors at Oklahoma State University were asked to use as many of the file questions as deemed appropriate for their second regularly scheduled hour-long examination. Six questions were chosen from the test file and included in the examination. (Appendix D). The examination was divided into split-half forms and was referred to as Tests 2A and 2B. Test 2A was administered to 330 students and test 2B was administered to 330 students. The questions used were chosen by the course instructor rather than the investigator, and there was no attempt made to control the average difficulty level, average discrimination index or taxonomical objective.

Since the number of questions from the test file used by the instructors for the second hour-long examination was only six, the instructors were again asked to use as many of the file questions as they deemed appropriate on their final examination in general chemistry for engineering or science majors. The instructors chose five of the file questions for inclusion in their final examination. (Appendix E). Again the examination was divided into split-half forms and was referred to as Tests 3A and 3B. The total number of questions used on the second hour-long examination and on the final examination were combined, and the confidence limits were calculated along with the calculation of the t-test for the significance of the test mean score.

Transferability of Test Items

This part of the study was designed to determine if observed test scores from a general chemistry course for engineering or physical science majors at Oklahoma State University could be used in conjunction with the final test file scores to establish a linear regression prediction line from which additional file scores can be predicted from observed scores. If evidence supported that this could be done, then correspondingly, observed test scores could be predicted from the test file for general chemistry students. The examination used to accomplish this purpose consisted of twenty questions extracted from the test file. (Appendix G). The twenty questions were divided into two groups of ten questions each. The first set was referred to as Test 4A and the second set was referred to as Test 4B. Tests 4A and 4B were designed to produce a difficulty level of 66.8%. The observed test item scores were used to establish a regression line with corresponding file item scores. A value corresponding to Test 4B observed values was determined from the regression line.

A t-test for significant differences was calculated to determine if these observed values varied significantly from their corresponding file values. Tests 4A and 4B each contained two questions from the knowledge category, two questions from the comprehension category and six questions from the application category. The tests were administered in the laboratory sections of general chemistry for engineering or physical science majors at Oklahoma State University. The papers were hand graded by the investigator rather than being computer scored.

Reliability Evaluation Using Bloom's Taxonomical Objectives

The verification of the test item classifications using Bloom's taxonomical objectives is a second purpose of this study. In order to verify the questions according to Bloom's scheme for classifying questions, an instructor at the University of Science and Arts of Oklahoma, Jeanne Mather, classified randomly selected questions from the test file according to the major taxonomical objective of each. (Appendix H). Mrs. Mather holds an A.A. degree in Social Sciences, a B. S. in Education and has done considerable graduate work dealing with Bloom's taxonomical objectives.

A t-test was calculated for the classifications of the 25 questions selected by Mrs. Mather to determine if any significant differences existed between her classifications and the classifications in the test file. Students were not asked to participate in this part of the study since only instructors should be involved in selecting questions from the file to be used on examinations.

Instruments

I. I.B.M. Card. Each student was asked to fill out his answer on regular I.B.M. cards.

II. CHEMOMR Program. The test file and its development were extensively discussed in the previous sections along with the development of the individual tests to be used. None of this would have been possible without the aid of Oklahoma State University's I.B.M. 360 computer and their CHEMOMR grading program. It not only grades the test results but provides extensive item and test analysis on those results. The following is furnished as a courtesy of Oklahoma State University.

> CHEMOMR is a program designed to read special optically marked tab cards used in multiple choice testing. The program produces listings of test scores with identification, number right, number wrong, and total score for each respondent. It is designed to accept single answer multiple-choice, true-false, and yes-no question. The students may mark more than one response. From one to five forms of the test can be given; each having a unique key. All test forms must have the same number of questions ranging from one to fifty. A weighting factor and penalty factor is used to calculate adjusted scores.
>
> Output consists of:
>
> 1. A listing of section, student name, student number, student right, number wrong, and adjusted score. The students are listed alphabetically within theory section and lab section. Means are printed for each section.

2. A listing of student name, student number, number right, number wrong, and adjusted score. The students are sorted by adjusted score, all students are included in this listing. The mean for all students is printed.

3. A histogram showing the number of students having adjusted scores in each equally spaced interval from zero to the maximum possible score. The interval is supplied by the user.

4. An item analysis by test form giving, for each question:
   a. The number of responses for A.
   b. The number of responses for B.
   c. The number of responses for C.
   d. The number of responses for D.
   e. The number of responses for E.
   f. The percentage of the students taking this test form who responded correctly.
   g. The item discrimination index which is (t-b)/n, where t is the number of students in the top 27% of the class (as determined by total score) taking this test form who responded correctly, b is the number of students in the bottom 27% of the class (as determined by total score) taking this test form who responded correctly, and n is the number of students in 27% of the class.

Hypotheses

1. (a) There is no difference (at the 0.05 level of confidence) in the mean on questions designed to assess the learning of chemistry by a group of students and the mean on questions from a developed test file so long as course variables remain essentially constant.

(b) There is no difference (at the 0.05 level of confidence) in the mean values on questions to assess the learning of knowledge or the application of chemistry by a group of students and the predicted mean on questions from a developed file.

(c) There is no difference (at the 0.05 level of confidence) in the mean scores on questions 4B from the question file and their predicted mean scores from a test given to an experimental group based on 4A.

2. There is no difference (at the 0.05 level of confidence) in the mean scores on questions on form A and the mean scores on its split-half form B, having identical questions in a different sequence.

3. There is no difference (at the 0.05 level of confidence) in the mean scores and there is no difference in the correlation coefficient between the individual questions when the observed scores are greater than .90.

4. There is no difference (at the 0.05 level of confidence) in the cognitive skill index for the taxonomical objective on the file questions and the index as assessed by other instructors.

Summary

A test file composed of 500 multiple choice type questions was compiled for the purpose of extracting suitable questions for incorporation into the examinations used in the pilot study and principal study. All questions in the file contained three working distracters. The discrimination indexes of the questions varied between +.25 and +.90, and difficulty levels were between 25% and 90%.

A pilot study was conducted to gather evidence to either support or reject major hypothesis number one which determines the transferability of the test file. The pilot test of 30 questions was given to seven students at the University of Science and Arts of Oklahoma, and the test mean scores were calculated from the percent correct responses on each item. The first ten test item scores and their corresponding file scores were used to establish a linear prediction line from which file values corresponding to a second ten questions were predicted.

The principal study was conducted to establish the reliability of the questions in the test file. This was accomplished by comparing the test scores on individual questions from the test file with the observed performance scores on the split-half test forms. The principal study was also designed to determine if observed test scores from a general chemistry course for engineering or physical science majors at Oklahoma State University could be used in conjunction with the final test file scores to establish a linear regression prediction line from which additional file scores can be predicted from observed scores.

The final part of the principal study was to determine the validization of classifying questions according to Bloom's taxonomy. This was accomplished by having another instructor, Mrs. Jeanne Mather, randomly select questions from the test file and calculating a t-test value between Mrs. Mather's classifications and corresponding file classification values to determine if any significant differences exists.

CHAPTER IV

PRESENTATION AND ANALYSIS OF THE DATA

Introduction

This study was conducted to determine if student's test item scores in general chemistry for engineering or science majors could be scaled to produce the same mean difficulty level when the same items were given to a class of general chemistry for engineering and physical science majors. A second purpose was to check the reliability of question classification in the developed test file against Bloom's taxonomical classification of cognitive levels. This comparison was restricted to the knowledge, comprehension and application categories.

This chapter presents the statistical tests and subsequent analysis of the data from the pilot study and the principal study. Homogeneity and reliability of items and sets of items along with the principal and secondary hypotheses were statistically tested.

Statistical Techniques

The t-test for paired groups was the main parametric test used to establish the probability of agreement between means at the .05 level of significance and to determine the validity of the null hypothesis. Herzer states that there is no real need for the artificial limitation of the .01 or the .05 levels of significance.[1] In educational situations the .20 or some other level might be just as appropriate as the widely used .05 level. In accordance with generally accepted practice, the .05 level of significance was chosen as appropriate to determine the validity of the null hypothesis.

The reliability of the test file item scores was ascertained by calculating a t-distribution score on the test mean scores and by calculating the confidence limits on each test file item used. The correlation of the individual items was supported by calculating a simple correlation coefficient for the various split-half test forms. This test is a measure of the degree of linear association between two variables. This differs from the t-test since the t-test establishes the degree of agreement between the test mean scores.

Homogeneity of variance of test 2A, for example, is important because the means of two sets of data may be the same while the variances of the two sets of data differ significantly. For example, two tests were given to two groups of general chemistry students, one group being engineering and science majors and the other group being engineering or physical science majors. The standard deviation of these two groups was vastly different, thus, producing a high f-test value. Therefore, the two groups are not homogeneous, and running a t-test on the results would be meaningless.

Linear regression was chosen as the method of correlating the test file item data to the observed test data. After the linear regression

---

[1]Harry Baldwin Herzer III, "A Study of the Effects of Single Concept Loop Films Upon Laboratory Techniques When Used for Pre-Laboratory Instruction in the Introductory Organic Chemistry Laboratory" (unpub. Ed.D. dissertation, Oklahoma State University, 1970).

analysis was performed, coefficients of determination were calculated to determine the quality of fit achieved by the regression.

Findings of the Pilot Study

The pilot study was to ascertain if there was a significant difference in the test mean scores calculated from mean item scores. The examination (Appendix A), composed of 30 questions, was divided into three parts (tests 1A, 1B, and 1C), each part consisting of ten questions. The f-values for tests 1A, 1B and 1C, comparing their variances to each other, were 1.23, 1.67 and 2.06, respectively, and are less than the tabled value of 2. Thus, the groups of data were homogeneous. (Table V). The t-values were calculated to be 1.43, 1.34 and .33, respectively. These values are less than tabled values for 9 degrees of freedom, thus, there was no significant differences in the student mean scores at the .05 level. Therefore, different tests can be constructed utilizing different questions from the test file, and the test designer can feel confident that each will produce a mean test score with no significant difference.

Tests 1A, 1B, and 1C contained two questions from the knowledge category, two questions from the comprehension category, and six questions from the application category. Each test was designed to produce a mean score of 66.7%. The file test item averages are listed in Table VI.

The means of tests 1A, 1B and 1C were treated to determine if different questions can be given to different students and obtain results with no significant differences in test mean scores. The results of the

TABLE V

CORRELATION COEFFICIENT, T-TEST AND F-TEST
RESULTS FOR TESTS 1A, 1B AND 1C
(N = 7)

	Mean Scores	t-test	f-test	df
File Question 1A/ Test 1A	66.80 67.00	.03	1.16	9
Test 1A/ Test 1B	67.00 55.80	*1.43	2.06	9
File question 1A/ File question 1B	66.80 66.80	0.00	2.88	9
File question 1B/ Test 1B	66.80 55.80	1.20	5.13	9
File question 1C/ Test 1C	66.60 52.80	2.16	2.27	9
Test 1A/ Test 1C	67.00 52.80	*1.34	1.23	9
Test 1B/ Test 1C	55.80 52.80	* .33	1.67	9
File question 1A/ File question 1C	66.80 66.60	.02	2.13	9
			t (.05)= 2.262 f (.05)= 4.03	

* There was no significant differences in the mean scores at the .05 level when tests 1A, 1B and 1C were compared to each other.

TABLE VI

MEANS FOR TESTS 1A, 1B AND 1C GIVEN AT THE UNIVERSITY OF SCIENCE AND ARTS OF OKLAHOMA

	File Mean	Test Mean	Standard Deviation File	Standard Deviation Test
1A	66.80	67.00	21.65	20.14
1B	66.80	55.80	12.76	28.91
1C	66.60	52.80	14.83	22.36

calculated t-test values are recorded in Table V. The calculated t-test values were: Test 1A/1B = 1.43; Test 1A/1C = 1.34; Test 1B/1C = .33. These calculated values are not significant at the .05 level. The f-value was calculated and the groups of data were found to be homogeneous. The correlation coefficients were: File 1A/Test 1A = .71; File 1B/Test 1B = .28; File 1C/Test 1C = .48. Tests 1A, 1B and 1C were designed to produce equal mean scores. The t-test, thus, established that there was no significant difference in test mean scores at the .05 level. There was a positive correlation between test file questions 1A, 1B and 1C.

The successful prediction of the test item scores illustrates that the same questions can be used in general chemistry classes which differ in instructional approach and student characteristics. Table VII shows the slope of the line and y-intercept value used to establish a least-squares prediction line from the raw test scores so that the test mean scores could be analyzed. The investigator was very encouraged when, after least-squares analysis, observed item scores for test 1A were successfully used to predict the test file item scores for test 1B.

A regression line of the observed values for test 1A on the corresponding file values was used for prediction purposes. (Table VIII). The observed values were plotted on the ordinate and the file values were plotted on the abscissa for test 1A. (Figure 1). The corresponding file values for 1B and 1C were predicted by locating the observed value on the ordinate and reading the corresponding value on the abscissa. (Table IX).

TABLE VII

LEAST-SQUARES ANALYSIS OF FILE QUESTION VERSUS TEST 1A

File Question versus Test 1A		
	Slope	.26
	y-intercept	49.67

TABLE VIII

REGRESSION LINE ESTABLISHED USING LEAST-SQUARES
FOR THE PREDICTION OF THE FILE SCORES
ON TESTS 1B AND 1C

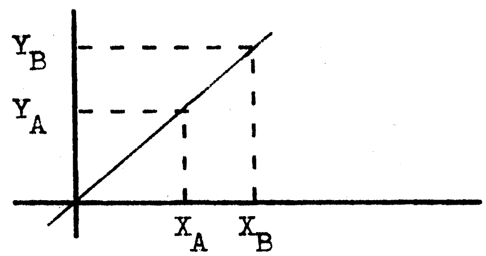

TABLE IX

T-TEST FOR SIGNIFICANCE OF PREDICTED VALUES
FROM THE OBSERVED SCORES AND SCALED
SCORES ON TESTS 1B AND 1C
(N = 10)

Scaled Scores	Observed Scores	df	Scaled Scores	Observed Scores	df
72.2	86	8	76.8	100	8
55	29	8	63.7	57	8
76.8	100	8	59.2	43	8
55	29	8	51.5	14	8
51.5	14	8	59.2	43	8
68	71	8	59.2	43	8
63.7	57	8	68	71	8
72.2	86	8	63.7	57	8
59.2	43	8	63.7	57	8

Table IX (Continued)

Scaled Scores	Observed Scores	df	Scaled Scores	Observed Scores	df
59.2	43	8	59.2	43	8

t (.05) = 2.306

The slope of prediction line produced from test 1A and its corresponding test file values had a slope of .26 and a y-intercept of 49.67.

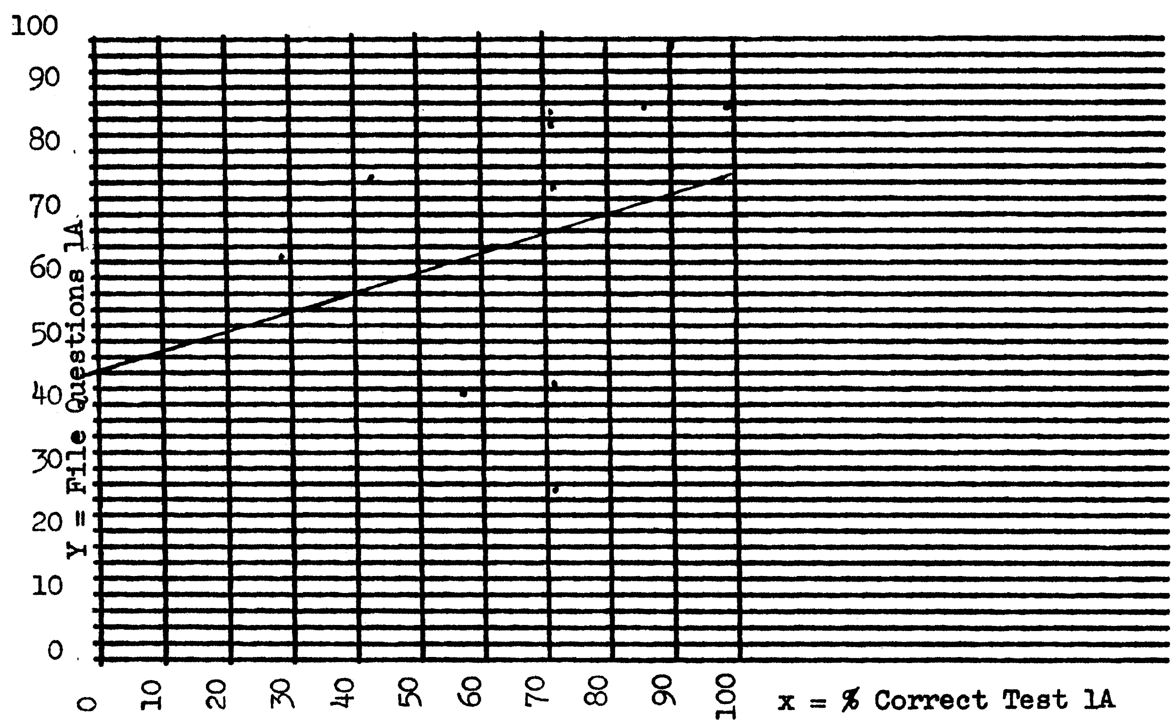

Figure 1. Linear Regression Prediction Line Formed by Plotting the Observed Difficulty Levels for Test 1A Versus Historical Difficulty Levels From File Questions

The t-values were calculated to be .38. The t-value calculated for test 1C had a value of 3. This exceeds the expected chart value of 2.262. The corresponding file values for test 1B were successfully predicted but the attempt to predict corresponding file values for 1C was not successful.

The last phase of the pilot study was to determine if a better correlation could be obtained if all questions with the same objective were grouped together. (Appendix C). Table X shows an analysis of questions grouped by objective.

TABLE X

ANALYSIS OF GROUPED TEST QUESTIONS AND FILE QUESTIONS

File Mean	Test Mean	Standard Deviation File	Standard Deviation Test	Corr. Coef.	Det. Coeff.	f-test	t-test
Objective 1 - Knowledge							
73.17	64.17	13.18	33.56	.29	.09	6.48	.68
				t (.05) 5 = 2.571		f (.05) 5 = 7.18	
Objective 2 - Comprehension							
71.67	61.83	12.85	19.37	.80	.63	2.27	2.00
				t (.05) 5 = 2.571		f (.05) 5 = 7.18	
Objective 3 - Application							
62.94	53.94	17.68	20.42	.06	.004	1.33	1.46
				t (.05) 17 = 2.11		f (.05) 17 = 2.89	

The correlation coefficients obtained by grouping and analyzing the items by objectives were no better. Two of the three correlation coefficients were low (.06 and .29), but all three t-values (.68, 2.00 and 1.46) support the assumption that there are no significant differences in the means of the questions when grouped according to objective. A correlation coefficient was determined for the mean scores of questions grouped according to knowledge, application, and comprehension. A value of 1.00 indicates a very strong linear relationship between the three levels of objectives. This seems to need further investigation in the principal study to see if this is caused by teaching strategies or by questions with lower objective numbers.

Findings of the Principal Study

This study was conducted to determine the limitations of transferability of carefully prepared test files and to determine if Bloom's taxonomy can be used by different instructors to place chemistry questions into the same classifications. The principal study was conducted during the first and second semesters of the fiscal year of 1976-77. All students enrolled at Oklahoma State University in general chemistry for engineering or science majors and three sections of general chemistry for engineering and physical science majors were involved in the study.

A set of 500 test file items was assembled. These items had been collected from tests previously given at Oklahoma State University in general chemistry classes for engineering or science majors. All test items had been used between 1973 and the spring semester of 1975. The instructors for the different sections of general chemistry for engineering or science majors selected various test file items which were judged appropriate for inclusion in the second hour examination. A copy of the test items used is included in Appendix D. Table XI contains a listing of the confidence limits calculated at the 95% level. The test file item with the lowest confidence limit was item 108 which was $\pm$ 28.68%. The best resulting confidence limit was calculated from test file item 210 which was $\pm$ 7.57%. The setting of these confidence limits helped to establish the validity of the test file items because accurate predictions cannot be made from scores with highly eratic confidence limits.

The validity of the test file mean item scores was further illustrated when the test file items used on the second hour examination were analyzed and the results listed in Table XII. The mean item scores were compared, and the confidence limit at the 95% confidence level was calculated to be $\pm$ 3.70%, thus, the validity of the questions was supported by both the t-test value and the high correlation coefficient for the split-half forms. The reliability of the test file item mean scores was established by analyzing the item mean scores from the second hour and final examinations. The reliability of the file values was supported by the t-test between the file value and split-half form A. The additional items chosen by the instructors for use on the final examination are listed in Appendix E. The analysis of the results are presented in Table XIII and XIV.

TABLE XI

CONFIDENCE LIMITS AT 95% LEVEL

File Number	Mean	Standard Deviation	Standard Error of Mean	Confidence Limit(95%)
106	40.3	4.04	2.33	10.02
107	75.0	4.36	2.52	10.84
108	46.7	11.60	6.67	28.68
176	68.3	3.1	1.76	7.57
178	78.7	5.9	3.38	14.53
210	56.3	3.06	1.76	7.57

TABLE XII

COMPARISON OF FILE AND SPLIT-HALF MEANS
TO PROVE RELIABILITY

	Mean	Standard Deviation	Standard Error of Mean	Mean Average	Confidence Limit(95%)
File	61.83	11.96	1.49	60.89	± 3.70
Split-half A	61.67	17.04			
Split-half B	59.17	19.49			

TABLE XIII

COMPARISON OF FILE AND SPLIT-HALF FORMS
TO PROVE THE RELIABILITY

	Correlation Coefficient	t-test	f-test	df
File/Split-half A	.79	.04	2.03	5
Split-half A/Split-half B	.98	1.50	1.31	5
		t = 2.571	f (.05)= 7.11	

TABLE XIV

COMPARISON OF FILE QUESTION MEANS AND
SPLIT-HALF MEANS ON FINAL
EXAMINATION

	Mean	Standard Deviation File	A	B
File Questions	64.4	15.63	15.79	13.31
Split-half A	61.80			
Split-half B	64.20			

	Correlation Coefficient	t-test	df	f-test
File/Split-half A	- .42	.18	4	1.06
Split-half A/ Split-half B	.96	1.09	4	1.41
		t (.05) = 2.776		
		f (.05) = 9.6		

Null hypothesis number one (a) which states there is no difference (at the .05 level of confidence) in the mean test file item scores of questions to assess the learning of chemistry by a group of students and the observed mean test item scores on the same question so long as course variables remain essentially constant was confirmed. An f-value of 2.03 and a t-value of .04 were calculated for the second hour examination, validating hypothesis number one. The f-test for variance illustrated that the internal variance in both sets of data had no significant differences, and the t-value demonstrated there was no significant difference in the mean item scores.

Null hypothesis number two which states there is no difference (at the .05 level of confidence) in the mean scores on form A and the mean scores on its equivalent split-half form B (having identical questions in a different sequence) was demonstrated to be valid by the calculation of an f-value of 1.41 and a t-value of 1.09 on the final examination. The f-test value illustrated that split-half forms A and B were homogeneous while the t-test value illustrated there was no difference in the mean item scores. This was further supported by the high correlation coefficients calculated for the split-half forms A and B on the second hour examination and final examination of .98 and .96, respectively.

Null hypothesis number three that there is no difference (at the .05 level of confidence) in the mean scores and the correlation coefficient between the individual questions when the observed scores were greater than .90 was not valid. The two correlation coefficients for the second hour examination and the final examination were .79 and -.42. Due to these low correlation coefficients hypothesis number three was demonstrated to be invalid. Neither correlation coefficient was high enough to warrant the making of individual decisions using individual questions.

The taxonomical objective classifications were demonstrated to be valid. The correlation coefficients correlating an instructor's classification of 25 questions with the identical questions in the test file were .96 and .97. (Table XV). These values are indicative of a very high correlation. The f-test produced a value of 1.03, thus, the two sets of classifications were homogeneous. A t-test value of .70 demonstrated there was no significant differences in the mean objective classification. Questions can be classified according to taxonomical objective with less than one chance in twenty of disagreement.

Student performance on questions designed to measure the acquisition of knowledge can be predicted from historical data at the 0.95 confidence level. This conclusion was supported by grouping test file questions used on the second hour examination and the final examinations by cognitive categories. Those questions classified into the knowledge category were analyzed together, and the questions classified in the application category were analyzed together. The t-test value of .68 and the f-test value of 1.25 reported in Table XVI clearly indicate that there is no significant difference between the mean scores on questions in the knowledge category and their corresponding test file mean. Therefore, hypothesis number one (b) is valid.

The objective of the test given in general chemistry for engineering or physical science majors was to establish a scale for the observed values that would allow the prediction of file values for other questions. The test was given at Oklahoma State University and consisted of 20 questions. (Appendix G). Questions 1-10 were referred to as test 4A,

TABLE XV

ANALYSIS OF BLOOM'S TAXONOMY
OF EDUCATIONAL OBJECTIVES

Expected Standard Deviation Sx	Observed Standard Deviation Sy	Correlation Coefficient	t-test	f-test	df
.96	.97	.80	.70	1.03	24
				t (.05) = 2.064	
				f (.05) = 2.39	

TABLE XVI

ANALYSIS OF KNOWLEDGE AND APPLICATION OBJECTIVES

	Knowledge	Application
File Mean	63.00	68.75
Split-half A Mean	53.33	68.00
Standard Deviation		
File	19.47	8.02
Split-half A	21.73	9.59
Correlation Coefficient	.13	.52
f-test	1.25	1.43
t-test	.68	.17
df	2	3
	t (.05) = 4.30 f (.05) = 39	t (.05) = 3.18 f (.05) = 15.60

and questions 11-20 were referred to as test 4B. The mean scores are reported in Table XVII, and the correlation coefficients, t-test and f-test are reported in Table XVIII.

Student performance in general chemistry for engineering or physical science majors cannot be obtained directly from historical data and be significant at the 0.95 confidence level. When the file item scores were compared to the observed results of test 4A (the first ten questions), a t-test value of 3.49 and f-test value of 1.89 were obtained. The sets of scores are homogeneous, but the t-test illustrates a significant difference in the mean scores at the .05 level. When file item scores were compared to the observed results of test 4B (the second ten questions), a t-test value of 3.34 was calculated. These two data sets are not homogeneous, thus, the t-test value was invalid. This demonstrated a difference in mean item scores which is significant at the .05 confidence level.

TABLE XVII

MEANS FROM TWENTY QUESTION TEST
GIVEN TO ENGINEERING MAJORS

	Mean	Standard Deviation
File/Test 4A	66.80	21.65
File/Test 4B	66.80	12.76
Observed Score A	43.70	29.79
Observed Score B	35.70	20.74

TABLE XVIII

ANALYSIS OF DATA FROM TWENTY QUESTION TEST GIVEN TO ENGINEERING MAJORS

	Mean	Correlation Coefficient	t-test	df	f-test
File Questions A/ Test 4A	66.80 43.70	-.71	3.49	9	1.89
Test 4A/ Test 4B	43.70 35.70	-.48	.58	9	2.06
File Questions A/ File Questions B	66.80 66.80	.28	0.00	9	2.88
File Questions B/ Test 4B	66.80 35.70	-.52	3.34	9	2.64
				f (.05) = 4.03 t (.05)=2.262	

A regression line was established using least-squares. This regression line had two purposes. One purpose was to determine the degree of linear relationship existing between the test file item scores and the observed scores on test 4A. The regression line was then used to predict the corresponding file values for the observed values of test 4B. The least-squares analysis of the test file item scores versus the observed scores on test 4A produced a slope of .52 and a y-intercept of 44.2.

Raw item scores from test 4A can be used to predict corresponding file values for test 4B from raw item scores. The predicted file item mean scores agree with the expected mean item score from the file at or above the 0.95 confidence level. The raw scores for test 4A were plotted against the file item scores and a regression line was drawn. (Figure 2). The file item scores for test 4B were predicted from the regression line, and the expected file values were predicted.

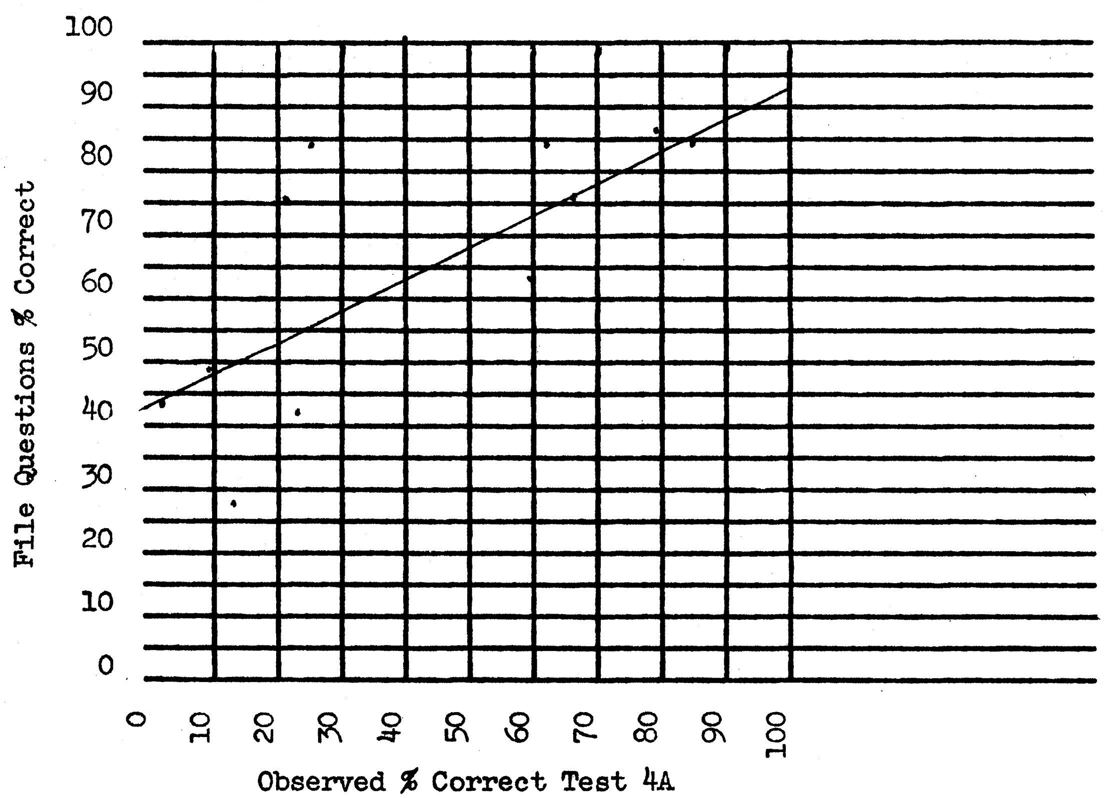

Figure 2. Raw Scores for Test 4A Plotted Against File Item Scores

After the raw file scores were predicted, there was found to be no significant difference at the .05 level between the mean score of the file items and the predicted file mean scores. Mean scores are reported in Table XIX. This is supported by an f-test value of 1.56 and t-test value of .28. The predicted mean scores of test 4B were compared to the historical performance (file scores) of the same question. The f-test proved that the sets of data were homogeneously grouped, and the t-test value illustrated there was no significant difference in test mean scores. Thus, there is no difference (at the .05 level of confidence) in the mean scores on questions from the question file and the predicted mean score from a test given to an experimental group (general chemistry for engineering or physical science majors) as stated in hypothesis one (c).

TABLE XIX

MEANS OF THE FILE QUESTIONS AND MEANS OF THE ADJUSTED TEST SCORES $4A^1$ AND $4B^1$

	Mean	Standard Deviation
Test $4A^1$	66.92	15.49
File Questions A	66.80	21.65
Test $4B^1$	67.04	6.81
File Questions B	66.80	12.76

TABLE XX

SCALED OBSERVED SCORES

4B	File Score	Observed Score
1	43%	2%
2	75%	21%
3	86%	62%
4	85%	85%
5	76%	66%
6	86%	79%
7	63%	60%
8	28%	13%
9	42%	23%
10	84%	26%

The following formula was used to establish the prediction line.

$$Y_1 = b\underline{\underline{x}}_1 + C$$

The slope of the prediction line (b) was .52. The y-intercept (c) was 44.2.

Summary

The confidence limit was set on the questions from the test file used on the second hour and final examinations in general chemistry for engineering or science majors at Oklahoma State University. The reliability of the test file items when used with general chemistry students with the same grouping characteristics was established. There was found to be no significant difference in the mean predicted scores on sets of ten questions when they were compared to their corresponding historical performance level.

The reliability of the mean score on sets of ten test file items was established by running an f-test to establish the homogeneity of the sets of data and a t-test to determine that there was no significant difference in mean test scores. This allowed student performance to be determined from historical data.

When the investigator's classification of questions according to taxonomical objectives was compared to another instructor's classification of the same questions according to taxonomical objectives, there was found to be no significant difference in the mean classification levels at the .05 level of significance. Thus, the testing for the mastery of cognitive skills can be accomplished while testing the student's level of achievement.

Student mean test scores in general chemistry classes, grouped according to characteristics other than those of students in general chemistry for engineering or science majors, were determined using the same sets of test questions. Before using the same test items in different classes the test item scores will have to be scaled for use in each specific type of general chemistry class.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

This study was conducted primarily to determine the transferability of test file items validated on a general chemistry class for engineering or science majors. This was accomplished by comparing historical mean scores from the test file to their predicted mean scores from observed raw scores. The instrument used to test the predictability of the mean scores was a 20 question examination where the first 10 and second 10 questions were referred to as tests 4A and 4B. To test for predictability the examination was given to a general chemistry class for engineering or physical science majors.

A pilot study was conducted to determine the value of conducting a larger scaled principal study. The pilot study consisted of one examination and was given at the University of Science and Arts of Oklahoma to determine that test file mean item scores could be used to predict mean test scores. The pilot study, thus, supported the results of the principal study.

The principal study consisted of three examinations. The first two were given in general chemistry classes for engineering or science majors. The purpose of these examinations was to establish that the test file items were reliable. These examinations were constructed by the regular course instructors, thus, only those file items deemed appropriate by these instructors were used.

The third examination was structured by the investigator and consisted of two sets of ten questions each. Each set of ten questions had the same mean level of difficulty and the same number of knowledge, comprehension, and application objective questions. This examination was administered to laboratory sections of general chemistry for engineering or physical science majors. It was administered by teaching assistants in charge of the laboratory sections. This examination demonstrated the classical Hawthorne effect which was observed and reported to the investigator by the instructor in charge of the course.

An instructor at the University of Science and Arts of Oklahoma classified 25 of the test file items according to taxonomical objective. These classifications proved to have a very high correlation to the classification of the items in the test file. The use of Bloom's taxonomy was validated.

A test file consisting of 500 questions developed by the investigator helped make this study possible. The questions were selected from criteria furnished by computer-analyzed item analysis. The item analysis was furnished to the investigator by Oklahoma State University and was selected on the basis of appropriate discrimination index, taxonomical objective, difficulty level and ability of the answers to discriminate.

Statistical tests used were the t-test for significance of the mean, f-test for homogeneity, correlation coefficient for the degree of linear fit, split-half forms for validity and confidence limit to determine item reliability. The null hypotheses were determined to be valid or invalid based on the f-test and t-test.

Conclusions

1. The ability to predict student performance from historical data was demonstrated in this study. The prediction of student performance was limited to the prediction of mean test scores. Individual test file item scores were reliable. Different series or sets of ten questions grouped by equal taxonomical objectives, discrimination indexes and difficulty levels were placed on tests and administered to students in general chemistry for engineering or science majors at Oklahoma State University. The mean score on the sets of ten questions were compared and found to contain no significant difference. Examinations can be constructed using different questions on each examination that produce equivalent mean scores. This allows for maximum individualization of examination, better instruction and better assessment of learning.

2. The assessment of learning is improved by the ability to individualize examinations and assess the degree of mastery of the cognitive levels of learning. For example, an instructor in history may desire his students to possess a knowledge of certain events taking place during the Civil War in the United States. A mathematics instructor may desire students be able to apply combination and probability theory in problem situations. The emphasis in these courses should be reflected by the objectives of the questions built into the test. This paper demonstrates that the degree of mastery of different cognitive skills can be evaluated by the same examination.

3. Test questions collected from previously given tests in general chemistry for engineering or science majors were given to general chemistry students majoring in engineering or physical science. The test file item scores on sets of ten questions were scaled to produce test mean scores in general chemistry for engineering or physical science majors that reflected no significant difference when compared to the mean score of the sets of test file items. This allowed the test file item scores to be scaled for use in general chemistry classes composed of students with different grouping characteristics and still produced a test mean score which was not significantly different from the test file mean item scores. The same test file item can be used with students at different universities but the results will have to be scaled before they can be interpreted. Since the file items can be used in various schools, their transferability has been established.

4. An instructor can classify questions according to Bloom's taxonomy and feel confident that the questions are placed into the correct category. An instructor at the University of Science and Arts of Oklahoma and the investigator classified the same questions and the scores had a correlation coefficient greater than 0.95.

5. A pilot study was conducted that demonstrated the value of the principal study. Examinations structured in the same manner as those used in the principal study were given to a class consisting of eight general chemistry students. The pilot study illustrated that student performance could be predicted from historical data and that raw score data from classes for engineering or physical science majors could be scaled so that test mean scores have no significant difference compared to the file test item mean scores.

6. The Hawthorne effect was observed in the test given to general chemistry students with engineering or physical science majors. The students were informed by the instructor that the test was part of a study and scores would not be used to determine class ranking. Student attitudes were reported by the instructor as being somewhat negative in taking the examination.

7. Accurate prediction of student performance cannot be made on the basis of one item score. This can be seen by looking at the confidence levels set on the second hour and final examination. The larger the number of questions used, the more valid will be the test mean score.

All the null hypotheses were verified except number three. The correlation coefficient for the raw scores and the test item scores did not correlate at the .90 level or above. Hypothesis number three was demonstrated to be invalid. The rest of the following null hypotheses listed below were verified:

There is no difference (at the .05 level of confidence) in the: test mean scores from historical performance on questions designed to assess the learning of chemistry by a group of students and the observed test mean score on questions from a developed test file so long as course variables remain essentially constant; test mean scores on questions to assess the learning of knowledge or the application of chemistry by a group of students and the predicted mean on questions from a developed file; test mean scores on questions 4B from the question file and their predicted mean scores from a test given to an experimental group based on 4A; mean scores on questions on form A and the mean scores on its split-half form B, having identical questions in a different sequence; and in the cognitive skill index for the taxonomical objective on the file questions and the index as assessed by other instructors.

Recommendations

This study explored some ways the power of the computer might be used to produce more meaningful examinations, which are important to the basic improvement of all teaching and specifically chemistry teaching. This study brings forth certain topics which demand further investigation.

1. Studies similar to this study should be conducted to determine why performance on some individual questions differ significantly from one semester to the next.

2. A study similar to this one should be conducted to determine what classes can adapt itself to the developed test file.

3. The percentages of knowledge, comprehension and application questions to be used on examinations in specific courses should be investigated.

4. The minimum number of questions from each taxonomical classification to produce a valid examination should be studied (as a function of the length of the test).

5. A study similar to this one should be conducted but using questions with higher discrimination indexes and a larger number of questions.

6. A study should be conducted to see if the difference in instructors will cause a variation in the mean scores.

A SELECTED BIBLIOGRAPHY

Alpert, D. and D. L. Bitzer. "Advances in Computer-Based Education." Science, 167 (1970), pp. 1582-1583.

Ansfield, Paul. "A User Oriented Computing Procedure for Compiling and Generating Examinations." Educational Technology, (March, 1973), pp. 12-13.

Avner, R. A., J. R. Ghesquiere, S. G. Smith. "The Use of Computers in the Teaching of Chemistry: Evaluation of Lesson Effectiveness." Journal of Chemical Education, 51 (1974), pp. 243-246.

Baker, Justine. The Computer in the School. Bloomington: The Phi Delta Kappa Educational Foundation.

Bloom, Benjamin S. et al. Taxonomy of Educational Objectives. New York: David McKay Company, 1956.

Brott, Eugene Charles. "A Study of the Effect of Selected Multiple-Choice Items on Conceptual and Factual Orientation of Students in General Chemistry." (unpub. Ed.D. dissertation, Oklahoma State University, 1969).

Castleberry, S. J., G. H. Culp, J. J. Lagowski. "The Impact of Computer-Based Instructional Methods in General Chemistry." Journal of Chemical Education, 50 (1973), pp. 469-472.

Christopher, Ronald. "Planting CAI." Educational Technology, (August, 1974), pp. 59-62.

Cochran, William G. and George W. Snedecor. Statistical Methods. 6th Ed. Ames: The Iowa State University Press, 1967.

Cromack, Theodore R. and Gilbert Sax. "The Effects of Various Forms of Item Arrangement on Test Performance." Journal of Educational Measurement, 3 (1966), pp. 309-311.

Cronbach, Lee J. "Course Improvement Through Evaluation." Teachers College Record, 64 (1963), pp. 670-673.

Ebel, Robert L. Essentials of Educational Measurement. Englewood Cliffs: Prentice-Hall, Inc., 1972.

Fast, Kenneth V. "An Analysis and Classification of the ACS-NSTA High School Chemistry Achievement Tests Using Bloom's Taxonomy-Cognitive Domain." Science Education, Vol. 58 (1974), pp 17-21.

Freese, Frank. Elementary Forest Sampling. Agriculture Handbook No. 232 (December, 1962) U. S. Department of Agriculture.

Freese, Frank. Elementary Statistical Methods for Foresters. Agriculture Handbook 317 (January, 1967) U. S. Department of Agriculture.

French, James R. and Ryan, James J. "Long-Term Grade Predictions for Intelligence and Achievement Tests in Schools of Differing Socio-Economic Levels." Educational and Psychological Measurement, Vol. 36 (1976), pp. 553-559.

Friedman, Morton. "Teachers' Cognitive Emphasis and Pupil Achievement." Educational Research Quarterly, Vol. 2, No. 1 (Spring, 1977), pp. 42-47.

Geisinger, Robert W. "Construct Validation of Hierarchical Taxonomies of Educational Objectives." Educational Technology, (May, 1973), pp. 26-29.

Goolsby, Thomas M., Jr. "The Validity of the College Level Examinations Programs Tests for use at the College Sophomore Level." Educational and Psychological Measurement, Vol. 30 (1970), pp. 375-380.

Hammond, K. R. and James E. Householder. Introduction to the Statistical Method. New York: Knopf Publications, 1963.

Hedges, William D. "How to Construct a Good Multiple-Choice Test." The Clearing House (September, 1964), pp. 8-11.

Hedges, William D. and Derrick Mark Pauw. "Computer Generated Tests for Independent Study Programs." NASSP Bulletin, (November, 1971), pp. 91-98.

Heinje, Cynthia L. "Junior College Computerized Instructional Support Services." Educational Technology, (August, 1975), pp. 32-34.

Herzer III, Harry Baldwin. "A Study of the Effects of Single Concept Loop Films Upon Laboratory Techniques When Used for Pre-Laboratory Instruction in the Introductory Organic Chemistry Laboratory." (Unpub. Ed.D. dissertation, Oklahoma State University, 1970).

Kerlinger, Fred N. Foundations of Behavorial Research. New York: Holt, Rinehart, Winston, Inc., 1964.

Kleinke, David. "A Linear-Prediction Approach to Developing Test Norms Based on Matrix-Sampling." Educational and Psychological Measurement, Vol. 32 (1972), pp. 75-84.

Knapp, Robert, et al. "The Validity of the Tests of Achievement in Basic Skills for Predicting Achievement in General Mathematics and Algebra." Educational and Psychological Measurement, Vol. 30 (1970), pp. 951-954.

Kropp, Russell P. et al. "The Construction and Validation of Tests of the Cognitive Processes as Described in the 'Taxonomy of Educational Objectives.' " Tallahassee: Florida State University, ERIC Report Resume ED 010-044, 1966.

Lehmann, Irvin J. and Mehrens, William A. Standardized Tests in Education. New York: Holt, Rinehart and Winston, Inc., 1969.

Lessinger, Leon M. "Test Building and Test Banks Through the Use of the Taxonomy of Educational Objectives." California Journal of Educational Research, XIV, No. 5 (November, 1963) pp. 195-201.

Libaw, Frieda B. "Constructing Tests with the Mentrey Tutorial Testing System." Educational Technology, 13 (1973), pp. 30-31.

Lindvall, C. M. Measuring Pupil Achievement and Aptitude. New York: Harcourt, Brace and World, Inc., 1967.

Lippey, Gerald. "The Computer Can Support Test Construction in a Variety of Ways." Educational Technology, (March, 1973), pp. 9-12.

Lohmann, Irvin and William Mehrens. Standardized Tests in Education. New York: 1969.

Molnar, Andrew R. "Viable Goals for New Educational Technology Efforts in Science Education." Educational Technology, (September, 1975), pp. 16-21.

Olympia, Jr., P. L. "Computer Generation of Truly Repeatable Examinations." Educational Technology, (June, 1975), pp. 53-55.

Payne, David A. The Assessment of Learning, Cognitive and Affective. D. C. Heath and Co., Lexington, 1974.

Remondini, David L. "Test Item System: A Method of Computer Assisted Test Assembly." Educational Technology, (March, 1973), pp. 35-37.

Salisbury, Alan B. "Computers and Education: Toward Agreement on Terminology." Educational Technology, (September, 1971), pp. 35-39.

Schonberger, Richard J. "Modular Instruction with Computer Assembled Repeatable Exams: Second Generation." Educational Technology, (February, 1975), pp. 36-38.

Storey, Arthur G. The Measurement of Classroom Learning. Science Research Associates, Inc., 1970.

"TIPS, a diagnostic tool to individualize instruction in the large class."

Tuckman, Bruce W. Measuring Educational Outcomes. New York: Harcourt Brace Jovanovich, Inc., 1975.

Van Hees, E. J. "Computer Managed Learning at the University Level in the Netherlands." Educational Technology, (April, 1976), pp. 28-31.

Wasik, John L. "Prediction of Success in a Behavioral Science Course." Improving College and University Teaching, Vol. XXII (Autumn, 1974), pp. 222-224.

Wood, Dorothy Adkins. Test Construction. Columbus: Charles E. Merrill Books, Inc., 1960.

APPENDIX A

THIRTY QUESTION EXAMINATION
(PILOT STUDY)

CHEMISTRY 1124 EXAMINATION
FALL, 1976

NAME________________________________

_____1. A water solution of pure zinc nitrate has a nitrate ion concentration of 0.1 moles/liter. How many gram formula weights per liter of zinc nitrate are present?

A. 0.1 B. 0.2 C. 0.3 D. 0.05

_____2. The concentration of Ba^{++} in a saturated aqueous solution of $BaSO_4$ is ($K_{sp} = 1 \times 10^{-10}$)

A. 1×10^{-10} B. 1×10^{-5}

C. 1×10^{-3} D. 1×10^{-8}

_____3. The quantity pH is defined by pH =

A. $-\log [H^+]$ B. $\log [H^+]$

C. $K_w / [OH^-]$ D. none of the others

_____4. Reactions which occur spontaneously always have

A. $\Delta H < 0$ B. $\Delta S > 0$ C. $\Delta G < 0$ D. $\Delta T = 0$

_____5. If a chemical reaction actually occurs, we know that for this reaction

A. ΔG is positive. C. ΔH is positive.
B. ΔG is negative. D. ΔH is negative.

_____6. At true thermodynamic equilibrium for a chemical or physical process, always

A. ΔG = negative C. ΔH = negative
B. $\Delta G = 0$ D. $\Delta S = 0$

_____7. If ΔG^o for the reaction: $C + \frac{1}{2} O_2 \rightarrow CO$ is -33 kcal and that for: $CO + \frac{1}{2} O_2 \rightarrow CO_2$ is - 61 kcal, what is the value of ΔG^o for $C + O_2 \rightarrow CO_2$?

A. +28 kcal B. -28 kcal C. -94 kcal

D. None of preceding is correct.

_____8. A 0.010 M solution of a monoprotic acid HX is observed to have a pH of 4. The equilibrium constant K_a for dissociation of HX is approximately

A. 1×10^{-8} C. 1×10^{-2}

B. 1×10^{-6} D. nonsense, since HX is completely dissociated.

_____9. One liter of an aqueous solution containing 0.10 moles of ammonium hydroxide and 0.20 moles of ammonium acetate had an OH^- concentration of approximately 1×10^{-5}. If 1 ml of 1.0 M HCl is added to this solution the pH will be approximately

A. 5 B. 9 C. 3 D. none of these

_____10. Given $K_w = 1.0 \times 10^{-14}$ for water at 25°C, what is the molar OH^- concentration of a 0.010 F HCl solution?

A. 1.0×10^{-12} B. 1.0×10^{-7}

C. 1.0×10^{-14} D. 1.0×10^{-16}

_____11. If $K_{sp} = 1.0 \times 10^{-4}$ for $CaSO_3(s) \rightleftarrows Ca^{+2}(aq) + SO_3^{-2}(aq)$ what is the SO_3^{-2} ion concentration when there is 2.0×10^{-3} moles/liter Ca^{+2} in solution?

A. 5×10^{-2} B. 5×10^{-1} C. 2×10^{-3} D. 1.0×10^{-4}

_____12. If s = solubility of $Fe(OH)_3$ in moles/liter, the expression for the K_{sp} is

A. $(s)(s)^3$ B. $(s)^3(3s)$

C. $\frac{1}{(s)(s)^3}$ D. $(s)(3s)^3$

_____13. At 25°C the value for K_w, the equilibrium constant for ionization of water, is

A. - 14 B. 14 C. 1.0×10^{-14} D. 1.0×10^{-7}

_____14. If you don't ignore the ionization of water, the pH of an aqueous solution of 1×10^{-10} M HCl would be calculated to be

A. greater than 9.
B. less than 3.
C. slightly less than 7.
D. slightly greater than 7.

_____15. For a spontaneous electrochemical reaction at constant temperature and pressure

A. Δ G must be positive.
B. Δ H must be positive.
C. E (the potential) must be negative.
D. E (the potential) must be positive.

_____16. To prepare a buffer solution having a pH of approximately 3, you would select a weak acid and a salt of that acid. The equilibrium constant for the ionization of the weak acid should be approximately

A. 3.0   B. 11.0   C. 10^{-3} D. 10^{-11}

_____17. The hydrogen ion concentration (moles/l) in aqueous 1.0 M HCN is: (K_a for HCN = 1×10^{-10})

A. 1×10^{-9}
B. 1×10^{-14}
C. 1×10^{-5}
D. 1.0

_____18. A solution has a pH of 11. The concentration of OH^- is

A. 1×10^{-3} B. 1×10^{-11} C. 3   D. 11

_____19. In the reaction $C_2H_5NH_2 + H_3O^+ \rightleftarrows C_2H_5NH_3^+ + H_2O$,

$C_2H_5NH_3^+$ is a

A. Brönsted acid.
B. Brönsted base.
C. Arrhenius acid.
D. Arrhenius base.

_____20. The correct equation for finding the acid dissociation constant (K_a) for

$$HX + H_2O \rightleftarrows H_3O^+ + X^-$$

A. $\dfrac{[HX][H_2O]}{[H_2O^+][X^-]}$

B. $\dfrac{[HX]}{[H_3O^+][X^-]}$

C. $\dfrac{[H_3O^+][X^-]}{[HX][H_2O]}$

D. $\dfrac{[H_3O^+][X^-]}{[HX]}$

_____21. An increase in entropy denotes an increase in

A. order B. randomness C. periodicity D. bond energy

_____22. Which of the following combinations of ΔH and ΔS for a chemical reaction indicate that the reaction should be spontaneous at all temperatures?

A. ΔH,-; ΔS,- C. ΔH,+; ΔS,-
B. ΔH,-; ΔS,+ D. ΔH,+; ΔS,+

_____23. For the reaction: $C(s) + 2H_2(g) \rightarrow CH_4(g)$ at 298^oK

ΔG_f^o and ΔH_f^o have the values -12.1 and -17.9 kcal respectively; ΔS_f^o is ____kcal/^oK.

A. 0.10 B. -0.10 C. -0.02 D. 0.02

_____24. If s = solubility of $PbCl_2$ in moles/liter, the expression for K_{sp} for $PbCl_2$ is

A. $\frac{1}{(s)\ (2s)^2}$ B. $(s)(s)^2$ C. $(s)(2s)^2$ D. $(s)(2s)$

_____25. The correct form for finding the solubility product constant (K_{sp}) for $Mg_3(PO_4)_2\ (s) \rightleftarrows 3\ Mg^{+2}\ (aq) + 2\ PO_4^{-3}\ (aq)$ is

A. $[3\ Mg^{+2}]\ [2\ PO_4^{-3}]^2$ C. $[3\ Mg^{+2}]\ [2\ PO_4^{-3}]$
B. $\frac{[Mg^{+2}]^3\ [PO_4^{-3}]^2}{[Mg_3(PO_4)_2]}$ D. $[Mg^{+2}]^3\ [PO_4^{-3}]^2$

_____26. K_{sp} for $AgBrO_3$ in H_2O is 6×10^{-5}. If you slowly add a solution of Ag^+ to a 0.001 M $KBrO_3$ solution, at what point would $AgBrO_3$ begin to precipitate?

A. when $[Ag^+] = [BrO_3^-]$ C. when $[Ag^+]\ [BrO_3^-] < K_{sp}$
B. when $[Ag^+] = K_{sp}$ D. when $[Ag^+]\ [BrO_3^-] > K_{sp}$

____27. A solution has a pH of 10. The concentration of OH^- is

A. 1×10^{-4} B. 1×10^{-10} C. 4 D. 10

____28. One property of an acidic solution is

A. pH > 7 B. $[H^+]$ less than 10^{-7} M

C. pOH > 7 D. $[OH^-] > [H^+]$

____29. A test for presence of Ba^{++} in solution is the formation of a white precipitate when SO_4 is added to the solution. What might you logically do to insure precipitation of as much $BaSO_4$ as possible?

A. Buffer solution to pH of 14.

B. Add considerable excess of $SO_4^{=}$.

C. Add considerable excess of $NO_3^{=}$.

D. Make sure solution is at pH = 7.

____30. Calculate the pOH of a 1.0 molar hydrochloric acid solution.

A. 1 B. 13 C. 0 D. 14

APPENDIX B

ITEM ANALYSIS OF TESTS 1A, 1B AND 1C
(PILOT STUDY)

ITEM ANALYSIS OF TESTS 1A, 1B AND 1C
(PILOT STUDY)

	File Question	Taxonomical Objective	Disc. Index	Percent Correct From File	Percent Correct From Test
Test 1A					
	402	3	.60	43	71
	398	3	.42	75	71
	431	1	.37	86	100
	383	2	.29	85	71
	380	1	.43	76	43
	382	2	.28	86	86
	372	3	.33	63	29
	428	3	.32	28	71
	417	3	.38	42	57
	447	3	.38	84	71
Test 1B					
	404	3	.60	59	86
	406	3	.48	71	29
	430	1	.37	80	100
	440	2	.43	53	29
	373	1	.45	72	14
	425	2	.49	62	71

File Question	Taxonomical Objective	Disc. Index	Percent Correct From File	Percent Correct From Test
424	3	.41	70	57
415	3	.38	76	86
418	3	.38	42	43
450	3	.36	83	43
Test 1C				
375	1	.49	77	100
378	2	.41	72	57
379	3	.38	51	43
400	3	.54	40	14
403	3	.32	82	43
407	3	.39	82	43
411	3	.69	72	71
414	1	.29	48	57
421	2	.41	72	57
446	3	.44	70	43

APPENDIX C

QUESTIONS GROUPED ACCORDING TO TAXONOMICAL OBJECTIVE (PILOT STUDY)

QUESTIONS GROUPED ACCORDING TO TAXONOMICAL OBJECTIVE

File Number	Taxonomical Objective	Percent Correct Question File	Percent Correct Test
430	1	80	100
373	1	72	14
431	1	86	71
380	1	76	43
375	1	77	100
414	1	48	57
- - -	- - -	- - -	- - -
440	2	53	29
425	2	62	71
383	2	85	71
382	2	86	86
378	2	72	57
421	2	72	57
- - -	- - -	- - -	- - -
404	3	59	86
406	3	71	29
424	3	70	57
415	3	76	86
418	3	42	43

File Number	Taxonomical Objective	Percent Correct Question File	Percent Correct Test
450	3	83	43
402	3	43	71
398	3	75	71
372	3	63	29
428	3	28	71
417	3	42	57
444	3	84	71
379	3	51	43
400	3	40	14
403	3	82	43
407	3	82	43
411	3	72	71
446	3	70	43

APPENDIX D

TEST FILE QUESTIONS INCLUDED IN SECOND
REGULARLY SCHEDULED HOUR-LONG
EXAMINATION
(PRINCIPAL STUDY)

TEST FILE QUESTIONS INCLUDED IN SECOND
REGULARLY SCHEDULED HOUR-LONG
EXAMINATION
(PRINCIPAL STUDY)

File #	% Correct File	% Correct Test 2A	% Correct Test 2B
106	41	44	36
107	70	77	78
108	60	40	40
176	71	69	65
178	72	81	83
210	57	59	53

QUESTIONS USED FROM TEST FILE

106. An example of a molecule with polar, covalent bonds is

A. NaCl B. H_2 C. HCl D. F_2

107. The difference between a covalent bond and a coordinate covalent (C.C.) bond is that

A. a C.C. bond is actually electrovalent.
B. in a C.C. bond both shared electrons are provided by the same atom.
C. the C.C. bond is formed only by the transition metals, i. e., only with d orbitals.
D. C.C. bond formation does not fit into the "rule-of-eight."

108. The type of bonding found in NH_3 is

A. ionic
B. covalent
C. coordinate covalent
D. ionic and covalent

176. The correct formula for tin(IV) chlorate is

A. $Sn(ClO_3)_2$
B. $Sn(ClO_2)_2$
C. $Sn(ClO_3)_4$
D. $Sn(ClO_2)_4$

178. The correct name for NaH_2PO_4 is

A. sodium hydrogen phosphate
B. sodium dihydrogen phosphate
C. sodium dihydrogen phosphide
D. sodium hydrogen phosphide

210. The mass, in grams, of 5.60 liters of O_2 gas at STP is

A. 16.0
B. 8.00
C. 4.00
D. 32.0

APPENDIX E

TEST FILE QUESTIONS INCLUDED IN FINAL EXAMINATION (PRINCIPAL STUDY)

TEST FILE QUESTIONS INCLUDED
IN FINAL EXAMINATION
(PRINCIPAL STUDY)

File #	% Correct File	% Correct Test 3A	% Correct Test 3B
106	41	62	62
176	71	70	70
210	57	78	75
233	75	63	72
451	78	36	42

QUESTIONS INCLUDED FROM TEST FILE

106. An example of a molecule with polar, covalent bonds is

A. NaCl B. H_2 C. HCl D. F_2

176. The correct formula for tin(IV) chlorate is

A. $Sn(ClO_3)_2$ B. $Sn(ClO_2)_2$

C. $Sn(ClO_3)_4$ D. $Sn(ClO_2)_4$

210. The mass, in grams, of 5.60 liters of O_2 gas at STP is

A. 16.0 B. 8.00 C. 4.00 D. 32.0

233. If a solution contains 0.5 mole of solute in 500 g of solvent, it is properly described as a ______ solution.

A. 1.0 molar B. 2.5 molar C. 1.0 molal D. 2.5 molal

451. A correct relation involving pH, pOH and the ionization constant for water K_w is

A. $(pH) \times (pOH) = pK_w$

B. $pH = pOH/K_w$

C. $pH + pOH = pK_w$

D. $pH - pOH = pK_w$

APPENDIX F

ANALYSIS OF THE QUESTIONS FROM SECOND AND FINAL EXAMINATIONS BASED UPON OBJECTIVES
(PRINCIPAL STUDY)

ANALYSIS OF THE QUESTIONS FROM SECOND AND FINAL EXAMINATIONS BASED UPON OBJECTIVES

File Question	% Correct File	% Correct Split-half A	% Correct Split-half B	Objective
106	41	44	36	1
107	70	77	78	1
451	78	36	42	1
176	71	69	65	3
178	72	81	83	3
210	57	59	53	3
233	75	63	72	3

APPENDIX G

TWENTY QUESTION TEST FOR ENGINEERING
OR PHYSICAL SCIENCE MAJORS
(PRINCIPAL STUDY)

TWENTY QUESTION TEST FOR ENGINEERING OR PHYSICAL SCIENCE MAJORS (PRINCIPAL STUDY)

File Number	Percent Correct From File	Objective	Percent Correct From Test
Test 4A			
402	43	3	2
398	75	3	21
431	86	1	62
383	85	2	85
380	76	1	66
382	86	2	79
372	63	3	60
428	28	3	13
417	42	3	23
447	84	3	26
Test 4B			
404	59	3	83
406	71	3	47
430	80	1	34
440	53	2	32
373	72	1	21
425	62	2	23
424	70	3	34

File Number	Percent Correct From File	Objective	Percent Correct From Test
415	76	3	23
418	42	3	51
450	83	3	9

CHEMISTRY 1515 EXAMINATION

NAME_______________________________ SPRING, 1977

DIRECTIONS: Place the letter of the answer you have chosen for each question in the space provided at the left of each question number.

_____1. A water solution of pure zinc nitrate has a nitrate ion concentration of 0.1 moles/liter. How many gram formula weights per liter of zinc nitrate are present?

A. 0.1 B. 0.2 C. 0.3 D. 0.05

_____2. The concentration of Ba^{++} in a saturated aqueous solution of $BaSO_4$ is ($K_{sp} = 1 \times 10^{-10}$)

A. 1×10^{-10} B. 1×10^{-5}

C. 1×10^{-3} D. 1×10^{-8}

_____3. The quantity pH is defined by pH =

A. $- \log [H^+]$ B. $\log [H^+]$

C. $K_w / [OH^-]$ D. none of the others

_____4. Reactions which occur spontaneously always have

A. $\Delta H < 0$ B. $\Delta S > 0$ C. $\Delta G < 0$ D. $\Delta T = 0$

_____5. If a chemical reaction actually occurs, we know that for this reaction

A. ΔG is positive. C. ΔH is positive.
B. ΔG is negative. D. ΔH is negative.

_____6. At true thermodynamic equilibrium for a chemical or physical process, always

A. ΔG = negative C. ΔH = negative
B. $\Delta G = 0$ D. $\Delta S = 0$

_____7. If ΔG^o for the reaction: $C + 1/2\ O_2 \rightarrow CO$ is -33 kcal and that for: $CO + \frac{1}{2} O_2 \rightarrow CO_2$ is - 61 kcal, what is the value of ΔG^o for $C + O_2 \rightarrow CO_2$?

A. + 28 kcal B. - 28 kcal C. - 94 kcal
D. None of preceding is correct.

____8. A 0.010 M solution of a monoprotic acid HX is observed to have a pH of 4. The equilibrium constant K_a for dissociation of HX is approximately.

A. 1×10^{-8} C. 1×10^{-2}

B. 1×10^{-6} D. nonsense, since HX is completely dissociated.

____9. One liter of an aqueous solution containing 0.10 moles of ammonium hydroxide and 0.20 moles of ammonium acetate had an OH^- concentration of approximately 1×10^{-5}. If 1 ml of 1.0 M HCl is added to this solution the pH will be approximately

A. 5 B. 9 C. 3 D. none of these

____10. Given $K_w = 1.0 \times 10^{-14}$ for water at 25°C, what is the molar OH^- concentration of a 0.010 F HCl solution?

A. 1.0×10^{-12} B. 1.0×10^{-7}

C. 1.0×10^{-14} D. 1.0×10^{-16}

____11. If $K_{sp} = 1.0 \times 10^{-4}$ for $CaSO_3(s) \rightleftarrows Ca^{+2}(aq) + SO_3^{-2}(aq)$ what is the SO_3^{-2} ion concentration when there is 2.0×10^{-3} moles/liter Ca^{+2} in solution?

A. 5×10^{-2} B. 5×10^{-1} C. 2×10^{-3} D. 1.0×10^{-4}

____12. If s = solubility of $Fe(OH)_3$ in moles/liter, the expression for the K_{sp} is

A. $(s)(s)^3$ B. $(s)^3(3s)$

C. $\frac{1}{(s)(s)^3}$ D. $(s)(3s)^3$

____13. At 25°C the value for K_w, the equilibrium constant for ionization of water, is

A. - 14 B. 14 C. 1.0×10^{-14} D. 1.0×10^{-7}

____14. If you don't ignore the ionization of water, the pH of an aqueous solution of 1×10^{-10} M HCl would be calculated to be

A. greater than 9.
B. less than 3.
C. slightly less than 7.
D. slightly greater than 7.

____15. For a spontaneous electrochemical reaction at constant temperature and pressure

A. ΔG must be positive.
B. ΔH must be positive.
C. E (the potential) must be negative.
D. E (the potential) must be positive.

____16. To prepare a buffer solution having a pH of approximately 3, you would select a weak acid and a salt of that acid. The equilibrium constant for the ionization of the weak acid should be approximately

A. 3.0 B. 11.0 C. 10^{-3} D. 10^{-11}

____17. The hydrogen ion concentration (moles/l) in aqueous 1.0 M HCN is: (K_a for HCN = 1×10^{-10})

A. 1×10^{-9}
B. 1×10^{-14}
C. 1×10^{-5}
D. 1.0

____18. A solution has a pH of 11. The concentration of OH^- is

A. 1×10^{-3} B. 1×10^{-11} C. 3 D. 11

____19. In the reaction $C_2H_5NH_2 + H_3O^+ \rightleftharpoons C_2H_5NH_3^+ + H_2O$, $C_2H_5NH_3^+$ is a

A. Brönsted acid.
B. Brönsted base.
C. Arrhenius acid.
D. Arrhenius base.

____20. The correct equation for finding the acid dissociation constant (K_a) for $HX + H_2O \rightleftharpoons H_3O^+ + X^-$

A. $\frac{[HX][H_2O]}{[H_2O^+][X^-]}$

B. $\frac{[HX]}{[H_3O^+][X^-]}$

C. $$\frac{[H_3O^+][X^-]}{[HX][H_2O]}$$

D. $$\frac{[H_3O^+][X^-]}{[HX]}$$

APPENDIX H

VERIFICATION OF BLOOM'S TAXONOMICAL OBJECTIVES

VERIFICATION OF BLOOM'S TAXONOMICAL OBJECTIVES

File No.	Type Learning	Observed Value	Expected Value
10	Application	3	3
23	Comprehension	2	2
58	Knowledge	1	1
87	Knowledge	1	1
111	Application	3	3
120	Application	3	2
148	Application	3	3
174	Knowledge	1	3
203	Application	3	3
243	Application	3	3
280	Analysis	4	4
286	Comprehension	2	1
306	Application	3	3
335	Knowledge	1	2
341	Knowledge	1	1
350	Knowledge	1	1
360	Application	3	3
396	Knowledge	1	1
403	Comprehension	2	3
407	Application	3	3

File No.	Type Learning	Observed Value	Expected Value
425	Comprehension	2	2
430	Comprehension	2	1
460	Application	3	3
485	Knowledge	1	1
490	Application	3	3

APPENDIX I

FORMULAS USED

FORMULAS USED

Coefficient of Determination

$$r^2 = \frac{\left[\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}\right]^2}{\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right]\left[\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}\right]}$$

Confidence Limits

$$L = \bar{X} \pm (t_{.05})(S\bar{D})$$

Correlation Coefficient

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

Mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Standard Deviation

$$S_x = \sqrt{\frac{\Sigma X_i^2 - n\bar{y}^2}{n-1}}$$

Standard Error of Mean

$$S_{\bar{D}} = \frac{S_D}{\sqrt{n}}$$

f-test

$$f = \frac{S_1^2}{S_2^2}$$

t-test

$$t = \frac{\bar{D}}{S_{\bar{D}}} \qquad t = \frac{m}{\sqrt{s^2 / \sum_{i=1}^{10} (X_i - \bar{X}_A)^2}}$$

VITA

Larry Maurice Scroggins

Candidate for the Degree of

Doctor of Education

Thesis: A STUDY OF COMPUTER-ANALYZED EXAMINATIONS

Major Field: Higher Education

Biographical:

Personal Data: Born in Stigler, Oklahoma, July 25, 1943, the son of Mr. and Mrs. Otis Scroggins.

Education: Graduated from Stigler High School, Stigler, Oklahoma, in May, 1961; received Bachelor of Science in Education degree from Central State University in 1966; received Master of Teaching degree in Education from Central State University in 1968; received Specialist in Education degree in Chemistry from Pittsburg State University in 1971; completed requirements for the Doctor of Education degree at Oklahoma State University in December, 1977. Member of American Chemical Society, Oklahoma Academy of Science, Oklahoma Education Alumni Council, State Board of Engineering for Professional Land Surveyors.

Professional Experience: Science Instructor, Oklahoma City Public Schools, 1966-67; Science Instructor, North Kansas City Public Schools, 1968-70; Chemistry Instructor, Neosho County Community Junior College, 1970-74; Graduate Teaching Assistant, Oklahoma State University, College of Arts and Sciences, 1974-75; Chemistry Instructor, Langston University, 1975-76; Chemistry Instructor, University of Science and Arts of Oklahoma, 1976-77.