THE UNIVERSITY OF ALBERTA

COMPUTERIZED TEST ITEM BANKING: FEATURES

by

MICHAEL DAVID CARBONARO

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1988

ISBN 0-315-42782-5

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Michael David Carbonaro

TITLE OF THESIS: Computerized Test Item Banking: Features

DEGREE: Master of Education

YEAR THIS DEGREE GRANTED: 1988

(Signed) Michael Carbonaro

PERMANENT ADDRESS:

214 - Ramona Ave

Palo Alto, California

94301

Date 16/1/88

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Computerized Test Item Banking: Features submitted by Michael David Carbonaro in partial fulfillment of the requirements for the degree of Master of Education.

....
Supervisor

....

....

Date Dec. 4, 1987

To my Mom, Dad, Nicholas and Concetta.

They make my achievements worthwhile.

Abstract

This study examines the major features associated with computerized test item banking. After a thorough review of literature, discussions with teachers, and a hands-on review of locally available computerized test item banks, 55 item banking features were identified. A survey questionnaire was developed that organized these 55 item banking features using the Wisconsin model which has the following three dimensions: *item structure*, *systemization*, and *automation*. Using the questionnaire, a survey was conducted in four Alberta school districts. The sample was comprised of 350 high school teachers mainly teaching in academic subject areas. A response rate of 56.85% was obtained.

The general consensus of those high school teachers who responded to the survey was that a computerized test item bank would be a very useful tool for assisting in the test construction process. Furthermore, a number of computerized test item bank features were identified as being important for high school teachers regardless of the subject they taught. High school teachers in the Sciences, Mathematics, English, Social Studies and Non-academic areas differed significantly on certain computerized test item bank features.

The 55 questionnaire items pertaining directly to features of computerized test item bank were analyzed using factor analysis techniques. This resulted in an underlying structure to these features consisting of five factors which closely resembles the dimensions identified by the Wisconsin model.

Recommendations are made for the direction future research in the area of computerized test item banking should take.

Acknowledgements

Throughout the course of this study I received valuable assistance and personal support from many individuals. I am particularly grateful to Dr. Gene Romaniuk for his technical assistance and personal concern for myself and the study. His expectations for maintaining steady progress and producing quality work have been driving forces for me. I have benefited from his candid comments and his willingness to discuss the many different facets of this project.

A sincere appreciation is expressed to Dr. Steve Hunka and Dr. Milt Petruk for their guidance and support of this study. Being able to tap Dr. Hunka's knowledge in the area of statistical analysis helped me considerably. Furthermore, as a graduate student at DERS I was fortunate enough to be touched by his enthusiasm for learning and the research process. Interacting with Dr. Petruk has always been a stimulating experience. His frequent expressions of approval for my ideas and faith in my abilities provided unending sources of motivation.

I would like to thank the many individuals of the participating schools for providing the time and energy in helping me to distribute and complete the questionnaires. Considering the present demands on schools and the time frame within which I was working, these individuals were extremely supportive.

The people who work at DERS, especially John Nesbit, Henry Schulz, Shawn Abbott and Marc Mandin, made the time I spent there an enjoyable and memorable experience. I thank them for their moral support and constructive suggestions. I am grateful to George Buck for sharing his knowledge on early teaching technologies. As well, many thanks to Dr. Dave Wangler. I could not have asked for a better friend.

I am deeply indebted to Norma Nocente and her family for their support and encouragement. They accepted me as part of their family and for this I will be forever grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE PROBLEM

> Classroom tests play a central role in the evaluation of pupil learning. . . . They provide relevant measures of many important learning outcomes and indirect evidence concerning others. . . . The validity of information they provide, however, depends on the care that goes into the planning and preparation of the tests. (Gronlund, 1985, p. 117)

1. Introduction

Gronlund (1985) suggests that in order for valid evaluation to occur, high quality tests must be constructed. As many teachers will attest, quality exams take many hours to develop. According to Hambleton, Anderson and Murray (1983) "it is not uncommon to hear teachers complain about the difficulties that they face in preparing and scoring classroom tests. These complaints often center on their own limited test development skills, on the poor facilities for producing tests, and on the amount of time that it takes to do the work properly" (p. 65). Teachers give numerous types of tests (quizes, unit exams and final exams) throughout the school year and often find it difficult to allocate sufficient time for the development of the tests. Furthermore, teachers often find themselves in a predicament because evaluation is also used as a means to hold them accountable for the curriculum (Stiggins & Bridgeford, 1986). As a result teachers need other resources to assist them in developing tests of high quality.

Over the past several years, two major topics of attention at the Province of Alberta, Department of Education (Alberta News Release, 1980 to 1987) have been *evaluation* and the return of the diploma exams for the academic subject areas, and the use of *computers in the school*. These two topics are not altogether unrelated, since there has been an effort to improve the quality of examinations through the use of computer technology. Romaniuk (1980) investigated the possibility of implementing a computerized item banking system for use in Alberta schools. At that time, however, the only established item banking systems were those which ran on mainframe computing systems, usually administered by Provincial or State agencies. Of course mainframe computer systems required a large capital investment to purchase and maintain. In his conclusion Romaniuk cited a study conducted for Alberta Education by Brown, Anderson and Louth (1980) which stated that teachers would welcome assistance in test construction by means of a computerized item bank. Romaniuk outlined a number of factors that would have to be considered before a mainframe computerized test item bank could be implemented in the Alberta educational environment. Two of these factors were: 1) a determination of the target population to which the system was meant to help, and 2) a recognition that technology is constantly changing and new more sophisticated hardware and software systems may provide other alternatives.

Technological change and the availability of new alternative computer systems has been clearly evident in Alberta schools. With the rapid proliferation of microcomputers in Alberta schools, an alternative to mainframe-based computerized item banks which were investigated by Romaniuk (1980), to microcomputer-based item banks has become available. In 1981, Alberta Education announced its commitment to implementing microcomputers into Alberta schools (Alberta News Release, 1981). Petruk (1986) indicated that the projected number of microcomputers in Alberta schools, as of January 1987, was nearly 27,000. Alberta ranks fifth in North America with regard to the number of students per computer. This dramatic increase in the number of computers in Alberta schools has lead to an increase in computer literacy among teachers. For example, Petruk (1986) indicates that there is heavy use of microcomputers for administrative tasks such as word processing, spread sheets, and data bases. Another administrative use, which he did not include is his survey, could be the use of microcomputers for a school–based test item banking system.

In other jurisdictions, the rapidly expanding capabilities of the microcomputer have resulted in the increased use of microcomputers in the field of testing. According to Hsu (1986), "Using microcomputers to store preconstructed items and [printing] them as tests . . . is becoming very common. This application is feasible not just because of the availabilities of the massive storage space and fast retrieval capabilities of the computer, but also because of its word processing capability" (p. 17).

Gronlund (1985) identifies item banking by microcomputer to be an innovative technique for classroom teachers in the storing and recalling of test items. Gronlund's textbook is supplemented with a workbook by Kanter (1985) which provides programs written in BASIC for analyzing test results. Gronlund emphasizes that both item banking and item analysis by microcomputer can play an important role for the classroom teacher for enhancing the evaluation process.

2. The Problem and Research Questions

Although the number of microcomputers in Alberta schools is rapidly increasing and many test item banking software programs are becoming available, it appears that few teachers are using these systems to develop classroom tests. One explanation for the limited use of microcomputer test item banking programs by teachers may be that existing software does not meet the teachers' needs. Other possible reasons are that teachers do not know which programs are available, excessive software costs, the lack of appropriate computers and printers, and lack of knowledge about test item banks.

Development of microcomputer software for the educational environment is often done without adequate research into what is required by the individual who will use the software (Deck & Estes, 1984). Thus, many educational software programs are unsuitable for most schools (Dudley-Marling & Owston, 1987).

Estes and Arter (1984) state that test item banks can be more effective for the user if the designers of the bank know what features are needed by the users. To date there is little published data to indicate features which teachers desire in a microcomputer test item banking system. To adequately develop microcomputer software to assist teachers in improving the quality of classroom evaluation, item bank software developers must know what features teachers need.

The major purpose of this exploratory study is to determine what features classroom teachers regard as important in an ideal microcomputer test item banking program. Therefore, the research questions are:

(1) What are the major features which high school teachers regard as important for inclusion in a computer-based test item banking program?

(2) Do high school teachers from various subject areas respond differently with respect to the features regarded as important for inclusion in a computer-based test item banking program?

A computer-based test item bank which is designed on the basis of teachers' needs could potentially improve the quality of tests, save teachers time in constructing tests, and make use of microcomputer technology which is presently available in Alberta schools. Awareness of teachers' needs should increase the likelihood of a computerized test item bank being successfully implemented in the classroom.

3. Limitations

One major limitation of this study was the use of a questionnaire which structured the subjects' responses. A large portion of the questions were limited to a "Likert-scale-format", which forced the respondents to limit their responses to a four-point scale, ranging from "Most Important" to "Not Important". If none of the four scales were acceptable, the respondents could select "Don't Know". The general information questions on page one of the questionnaire (see Appendix B) were designed to allow for specific responses only.

The study was restricted to high school teachers in four urban city school districts in the Province of Alberta. A majority of the teachers who responded taught academic subjects. Thus the results are not generalizable to all high school teachers because the sample of teachers used in this study was not representative of the total population of high school teachers in the province of Alberta. Schools which participated in the study did so on a voluntary basis and were not chosen randomly.

4. Overview

Chapter I presented an introduction to the problem, purpose, significance and limitations of the study. Chapter II reviews various definitions of computerized test item banks, reasons why teachers test, and the problems with existing classroom tests. Some of the benefits of using computerized test item banks are described and a historical review of the application of computers in constructing tests via item banks is presented. In addition, Chapter II describes the main features of computerized test item banks and the necessary factors to consider when designing or choosing a microcomputer test item bank. Chapter III includes a description of how the questionnaire was developed, a description of the sample surveyed, and a description of the collection of data. Chapter IV presents the findings, including an analysis and interpretation of the data. Chapter V discusses the findings and makes recommendations for further research.

CHAPTER II

REVIEW OF THE LITERATURE

1. Introduction

The relationship between technology and testing has been a long and natural one (Baker, 1984). As early as the 1890s, Herman Hollerith developed an electronic device to tabulate the results of questions asked during the United States census (Augarten, 1985). In the mid 1920s Pressey (1924) developed a testing machine which stored questions and kept track of student errors. The Pressey machine allowed only multiple choice test items, and students would eventually have to get the question right before moving on. Later, Pressey (1925) revised his machine to include statistical information about the item distractors. In 1930, LaZerte, created a device called the "problem cylinder", which according to Buck (personal communication, June 1987) "was a primitive mechanical test bank". Test items were recorded on cards, and the mechanical device was capable of evaluating the right answer (Buck, in press). The teaching machine developed during the 1950s by B. F. Skinner incorporated the storing of test questions and a student evaluation method (Skinner, 1952). The early efforts of these individuals to use the available technology to lessen the workload created by the task of exam preparation and to improve the process of testing were not in vain. Recent applications of digital computers to assist teachers in the preparation of classroom tests are a logical progression from earlier efforts.

The idea of using computers to store and select test items is not new. Educators in the mid 1960s were using computers to store test items so that the items could be easily recalled (Stodola, 1965). At that time, items were punched on cards along with a code describing the instructional objective which they were designed to test. By selecting items based on this code, tests were constructed and were produced on mimeograph stencils. A mark-sense card reader was used to analyze the test responses. By the early 1970s, computers were successfully being used in test scoring, reporting of test results, and item analysis (Baker, 1971).

The application of computer technology to testing developed in two stages (Hsu & Sadock, 1985). The first stage being the 1960s and the first half of the 1970s, when large mainframe computers were used to assist instructors. The second, and most recent stage, began in the late 1970s and continues today; it primarily centers around using the ever increasing power of microcomputers (Baker, 1984, 1986; Broduer, 1986; Brzezinski, 1984; Hiscox, 1983; Nitko & Hsu, 1983; Peterson & Meister, 1983). As a result of the improved technology and increased availability of microcomputers, the application of computers to testing is becoming more widespread.

The March, 1973, issue of *Educational Technology* contained 22 articles related to Computer-Assisted Test Construction (CATC). Computers were recognized as a valuable tool in helping to construct a test. Gerald Lippey (1973), guest editor of that issue, began by saying that "assistance in the preparation of tests and exercises is a recently emerging use of computers that has very little publicity but shows signs of widespread appeal" (p. 10). Lippey's statement set the stage for the synthesis of many ideas associated with the application of computers to testing. He hoped the sharing of information on this topic would encourage the development of better systems because, until that time, many different groups had been working individually on projects concerning test development by computer (Lippey, 1974). Most of this early work of using computers to help in test development was being done at those colleges and universities that had the large computer facilities readily available (Seely & Willis, 1976).

In 1974, Lippey released the first book on the application of computers to assist in test development. The book, *Computer Assisted Test Construction,* focused solely on showing how computers could assist in the test construction process. According to Lippey, the book, ". . . was prepared primarily for educators who are interested in taking advantage of computers to aid in providing tests and exercises" (p. viii). The book has become the definitive work on the topic of Computer-Assisted Test Construction (CATC). The technological references in the book have become somewhat outdated, but the major outline Lippey followed is very cogent to recent work in this field. Lippey's outline included sections on item banking, item classification and selection, printing tests, the role of statistics, pedagogical implications and systems considerations.

Lippey's book defined CATC in terms of five functions that can be preformed by a computer. They are as follows: 1) *Item banking* — "refers to the storage of questions in a machine readable form" (p. 5), 2) *Item Generation* — items can be generated by the computer in accordance with a predefined algorithm, 3) *Item Attribute Banking* — in this case properties related to the item such as statistical data can be stored with the item, 4) *Item Selection* — items can be selected by their attributes, and 5) *Test Printing* — items retrieved from the bank are assembled for printing in some specified format. Lippey points out that there is overlap in these areas when developing a test by computer.

2. Problems Concerning an Operational Definition of CATC

Although Lippey (1974) defined CATC in terms of the above five functions, a problem exists in the literature concerning an operational definition of CATC. This problem stems from the fact that individual applications of computers to testing depend largely upon the *local area objectives* (Byrne, 1976; Johnson & Maher, 1982; Lippey, 1974; Millman & Arter, 1984; Newbold & Massey, 1977). According to Estes and Arter (1984) these local objectives reflect needs such as: student screening, classroom testing, competency testing, student diagnosis, mastery learning, survey assessment, and program evaluation. Therefore, the definition is dependent on the purpose behind which the system is developed.

The March, 1973, issue of *Educational Technology*, described 19 different banks, all with some unique characteristics. For example, the TCAP (an interactive computer program for test construction and analysis) system developed by Baker (1973) emphasized interactivity at the computer terminals and statistical item analysis. The MCQ (Multiple Choice Question) banking system was designed only for objective-based, multiple choice questions (Buckely-Sharp, 1973). MENTREX (Mentor or Examination) (Libaw, 1973) was designed for a commercial environment, and CTSS (Classroom Teacher Support System) (Toggenburger, 1973) was designed with the classroom teacher in mind. Consequently, because each of these systems emphasized a different aspect, a variation existed in the conceptual view of how computers should be applied to the testing process.

Several operational definitions for CATC have been developed (Muizniеks & Dennis, 1979). The most recent definition comes from Hsu and Sadock (1985) who define CATC as "any activity which involves the utilization of computer technology to construct items and/or to select a set of pre-constructed items to form a test" (p. 6). Although they narrow their definition by stating test construction should be based on "sound measurement principles" (p.6), it is more comprehensive than Lippey's definition because it allows for the inclusion of such features as on-line testing and adaptive testing (Millman, 1984).

Byrne (1976) redefined CATC in terms of question banks (Computerized Question Banking System [CQBS]) in which he states a "question bank is a simple collection of test questions organized to enable easy access to particular types of questions within the bank" (p. 45). In this case, Byrne's emphasis is on applying computer technology to enhance existing manual item banks. This seems sensible, considering most teachers were not familiar with computers, but were often familiar with the concept of item banks or question banks.

In reviewing the literature, Millman and Arter (1984) identified the question "What is an Item Bank?", as being one of several important issues in the field of computerized test item banks. More specifically, Byrne's idea of "question banks" and Lippey's (1974) term "item bank" differ in that Lippey restricts his definition to questions that are in a machine readable form. Newbold and Massey (1977) limit the term item bank to a "collection of *proven questions* [italics added] from which users draw a test conforming to their requirements" (p. 115). The emphasis being that all questions must reach a level of acceptance. Often this is done by pretesting the items or through a "shredding process", in which a team's content specialists review the questions, as a group, and determine if the question meets a specific criterion (Chelu & Elton, 1977; Johnson & Maher, 1982). Wright and Bell (1984) define an item bank as a "composition of coordinated questions that develop, define, and quantify a common theme, and thus provide an operational definition of a variable" (p. 331). The emphasis here is on the statistical properties involved with each question and probabilistic model for what might occur when a student attempts to answer a question. Burke, Kaufman and Webb (1985), while describing The Wisconsin Item Bank, defined the term item bank according to three variables which define a three-dimensional space.

Millman and Arter (1984) point out that Choppin (1976) made distinctions between, "question banks", "item pools", "item collections", "item reservoirs" and "test item libraries", whereas they have chosen to use these terms synonymously. On the other hand, the Portland Test Item Bank System makes specific distinctions between the terms "item pool" and "item bank" (Hathaway, Houser & Kingsbury, 1985). They refer to the term "item pool" as a specific set of items not having been "scaled to a common metric", while items that have been scaled to a common metric are part of an item bank (Hathaway, Houser, & Kingsbury, 1985, p. 32).

Estes and Arter (1984) provide a definition which seems quite general. They define item banking, "as a large collection of distinguishable test items" (p. 5). According to Estes and Arter, the focus of their definition is on three specific terms: large, collection, and distinguishable. *Large* refers to a test bank which contains more items than would normally be available for any given test. *Collection* means individual test items are gathered together for easy recall and this is usually the case in a computerized item bank. Finally, *distinguishable* means that the items contain specific attributes which provide a means by which the items can be separated. The item attributes "might be an assigned description of the item (e.g., content/objective classifications) or be a measured aspect of the item (e.g., difficulty level)" (p. 5).

In conclusion, there are various definitions of the term item bank and each definition is related to the circumstance for which it has been developed. For the purpose of this study, it seems necessary that the term item bank fits two general criteria: 1) that it is computerized, and, 2) that it is applicable to the classroom environment. As a result most of the above definitions have some degree of relevancy.

3. Why Are Tests Used?

Before introducing a computer system to assist in the testing process, "an educationally valuable purpose for such testing" (Fremer, 1984, p. 14) should be defined. Testing often forms the primary means of educational evaluation. Reports, such as The Nation at Risk (1983) and The Governors Task force on Education (1986), have received extensive media coverage and hence the return to prominence and use of many government (local, provincial, and federal) implemented competency based exams.

The public fear that the quality of our educational system is deteriorating has increased the demand for more stringent testing processes. Estes and Arter (1984) note that:

> Tests have long been the primary benchmarks by which effective schools or education are measured. Decreases or increases in test scores are often the most public measures of school and district effectiveness. The movement toward minimum competency requirements during the late seventies is a primary example of accountability in action. (p. 1)

Evaluation is not only a means by which the student's performance is judged, but that of teachers as well. In some educational jurisdictions, standardized tests are used, in part, to hold teachers accountable for the prescribed curriculum. As a result, it becomes important for teachers to understand the underlying principles of testing.

Individual classroom teachers usually use tests to ascertain (a) whether the student possesses the prerequisite knowledge level in a particular subject, (b) whether the student exhibits deficiencies during a learning activity, and (c) the level of learning which has taken place, e.g., mastery learning (Muiznieks and Dennis, 1979). In addition, Borg and Gall (1983) claim that testing evaluates how much students know and the mere presence of testing increases what they know. Therefore, care and attention must be given as to how individual tests are to be constructed.

4. Constructing Quality Tests

Constructing an effective test is a difficult and time consuming process. Nitko (1983) outlines the following test construction principles which should be addressed when attempting to develop a quality test:

1. Define the purpose for testing at this time.

2. Specify the performance and processes to be observed and tested.

3. Select the performance and processes to be used to observe and to test the performance.

4. Develop the initial drafts of the test exercises.

5. Are the items of satisfactory quality? If not, revise or reconstruct the items.

6. Do the items match the stated performance to be assessed? If they don't, revise or reconstruct the items.

7. Conduct a preliminary try-out of items, if possible.

8. Do the items appear to be functioning as intended? If they don't, revise or reconstruct the items.

9. Develop the final version of the test.

10. Administer the test and analyze the results.

11. Does the test appear to be functioning well? If it does not, revise or reconstruct the items.

12. Use the test for decision-making.

A properly constructed test, addressing the above guidelines, would provide the teacher with valuable information that can be used to improve the learning process. Many of Nitko's points focus on the quality of the items. Most teachers who have developed quality test items can testify to the work involved with item building. Often teachers with the best intentions are unable to develop a comprehensive examination which truly addresses the needs of the learner and the conceptual material covered in class (Zacharias, 1979; Davis and Zacharias, 1982). Although the time and effort is well spent, most teachers can not, nor should not, be expected to produce such a level of examination quality. For test items to reach a realistic level of respectability, more than one teacher must become involved in the item development. Thus, many hours of work are required and statistical item analysis needed to develop valid and reliable test items (Chelu & Elton, 1977; Gronlund, 1985). As a result, it must be emphasized that good quality test items, which have obtained a sufficient level of reliability and validity, are necessary for effective evaluation (Baker, 1973; Hambleton, 1984; Hathaway, Houser, & Kingbury, 1985; Hiscox, 1983; Wright & Bell, 1984). "There is no better way for a teacher or professor to continue to improve his skills in testing, and the quality of tests he uses, than to analyze systematically the results from tests and to compare the findings of these analyses with ideal standards of test quality" (Ebel, 1972: p. 93). According to Zacharias (1979) this may help to avoid many of the negative connotations that have long been associated with testing.

5. The Problem With Existing Classroom Testing

Classroom assessment provided by the teacher is important if education is said to be effective (Stiggins, Conklin & Bridgeford, 1986). Furthermore, the tests which "matter most to both students and teachers are classroom tests, created or chosen by teachers to mirror each segment of their curriculum as it is covered" (Heartel, 1986, p. 3). At least 10% to 15% of class instruction time is spent on testing (Haertel, 1986); similar results were reported in a recent review by Green and Stager (1986). Gullickson (1982) reported that the average amount of time spent on classroom testing is 20% or more, and that 98% of the 336 teachers he surveyed in South Dakota used tests on a bi-weekly basis. Such tests could be used to provide students with a means of determining what curriculum material is important in a course, in addition to finding out whether or not they know the curriculum material. With this large amount of time spent on classroom testing and with the importance these tests have for students in judging their progress, quality classroom testing becomes an important concern.

Ironically, a number of problems exist with classroom testing as it is presently applied in the school environment. Most teachers lack the proper training to assess children correctly; they do not have sufficient time nor the proper tools (Gullickson, 1985). McLean (1985) in a report to the *Canadian Education Association*, on policies, practices and uses of assessments of student achievement, concluded that most teachers know very little about classroom evaluation:

> Teachers receive little formal training in evaluation (sometimes none at all). They are seldom presented with systematic theoretical knowledge about evaluation, much less expected to master it. They make judgements about when evaluation techniques are to be applied, but the range of options is quite restricted. . . . Teachers learn about evaluation as potters learn about working with clay—from other skilled practitioners. Most know little about the underlying theories—why one technique works in a given situation and another does not. According to their accounts, they learn by experience with little or no supervision, and inservice training opportunities are becoming fewer and fewer every year. Unlike teaching in general, evaluation could often be scientific, but as it is generally practiced, we are far from a science of evaluation. Most teachers become skilled at evaluation-some are less skilled and some are professionals (in the sense defined above), but most apply a very few specialized techniques according to general rules that are rarely stated explicitly. They are made uncomfortable when they have to explain how they do it, and more uncomfortable still when asked how they justify doing it that way. (p. 33)

Similar concerns, have been found in a number of recent studies (Anderson, 1987; Carter 1984; Haertel, 1986; Green & Stager, 1986; Gullickson, 1982, 1984; Gullickson & Ellwein, 1985; Stiggins & Bridgeford, 1985). Haertel (1986) examined classroom testing and its role in classroom assessment by studying 1 900 students and 250 teachers in a number of San Francisco Bay area high schools. The findings indicated that "the content of most teachers' unit tests closely matched their instruction, but classroom examinations often failed to reflect teachers' stated instructional objectives, frequently requiring little more than repetition of material presented in textbook or in class, or solution of problems much like those encountered during instruction" (Haertel, 1986, p. 2). Carter (1984) studied 310 high school teachers and found that they had trouble writing test items which test higher level thinking skills. Similarly, Green and Stager (1986) studied Wyoming teachers and found that the use of short answer test items increased significantly as the grade level increased. Usually, such items are said to be associated with lower level cognitive skills (Gronlund, 1985).

Gullickson and Ellwein (1985), reported that few teachers ever compute the statistics necessary to evaluate test performance. Furthermore, teachers rarely ever carry out item analysis, nor do they consider it important in improving the quality of their tests. Fewer than 42% of the teachers computed simple range scores and only 10% to 13% use such information as the mean, median and standard deviation (Gullickson, 1982). Part of the reason for not using statistics may be that teachers consider their use too time consuming in opposition to their benefit (Gullickson, 1986). Of the teachers Green and Stager (1986) surveyed, 27% had no formal coursework in test theory, 47% had one course, 17% had two courses, and 9% had three or more courses. In addition, people with more coursework were reported to test on a more frequent basis.

Stiggins and Bridgeford (1985) point out that teachers at the high school level usually construct their own tests, rather than rely on those supplied by book publishers. As a result, these tests will vary from classroom to classroom and the data they provide is often unreliable. This is reinforced by McLean's claim that, "university dissatisfaction with school marks is based on a perception that standards vary greatly from school to school, and that there has been uneven inflation of marks" (p. 17). Furthermore, the fact that there is no standard evaluation criteria between classroom testing practices, prevents a fair and equal evaluation of the student.

Three-quarters of the 288 classroom teachers surveyed by Stiggins and Bridgeford (1985) indicated they were concerned about the assessment procedures they used. Yet, there seems to be little opportunity for teachers to improve their skills, particularly with school inservice programs being cut due to budget constraints (McLean, 1985). One way teachers can improve their evaluation skills is by working with colleagues who already have the skills. Research by Stiggins and Bridgeford (1982) indicates that teachers regard colleagues as one of the primary sources of evaluation ideas, unfortunately they rarely work together in developing test items and tests.

Because of the poor quality of current classroom testing practices, it is not surprising that one of McLean's (1985) final recommendations to the Canadian Education Association, regarding improvements to classroom student evaluation, includes the application of computer technology to the development of item banks. According to McLean (1985), "powerful microcomputers and microcomputer clusters are now appearing in schools . . . with enough storage capacity for the development of local item banks" (p 43). Furthermore, "the Calgary Board of Education is off to a fast start with a mathematics item pool based on a minicomputer in the board office. . . . the system gives teachers a look at what is possible and creates a nucleus of teachers with hands-on experience in using a computer-based system" (p. 43).

6. Benefits of Using a Computerized Test Item Bank

A number of reasons why computerized item banks should be used by classroom teachers (Baker, 1986; Broduer, 1986; Byrne, 1976; Fremer, 1984; Hiscox, 1983; Johnson & Maher, 1982; Lippey, 1974; Muiznieks & Dennis,1979; Nelson, 1984; Newbould & Massey, 1977; Nitko & Hsu, 1983; Nitko & Hsu, 1984; Nitko & Sadock, 1985; Ward, 1984; Wright & Bell, 1984) According to Millman and Arter (1984, p. 318), item banks have the greatest potential when one of following conditions exist:

1) The tests that are presently available do not suffice and new tests must be developed in accordance with local specifications

2) Testing occurs frequently.

3) Multiple forms of the same test are needed.

4) Individually tailored tests are required.

5) Multiple users collaborate in contributing test items to build a bank, and to create tests from the bank.

6) An item banking system (hardware and software) is available.

Millman and Arter's final point indicates that just having the available computer resources makes item banking potentially beneficial.

According to Johnson, Willis, Seely, and Moore (1981), a computerized test item bank can have a number of potential benefits. First, the computer can save the instructor time in creating a test. This time saving can be a result of the computer generating new test items, predetermined by an algorithm, judging the quality of existing items after a test is optically scored, and making editing of test items easier. Second, instruction can be more individualized and innovative. The computer can be used to generate a number of unique examinations and problem sets, all of approximately equivalent difficulty. This permits students to be tested more than once on a similar topic. It also provides instructors with an opportunity to design self-paced lessons in accordance with each student's needs by tailoring the difficulty of tests. Third, student performance can more easily be maintained and monitored if an optical scoring facility is available. Records pertaining to each student's test performance can be recorded and reported by the computer. This permits instructors to identify deficiencies and possibly provide students with remedial feedback. Finally, instructors can analyze student responses and thereby make continuous improvements to their test questions. "The overall quality of an item collection, properly modified as a result of experience, will continue to improve, even though the quality initially might leave much to be desired" (Lippey, 1974, p. 17). Information gained by means of item analysis can help identify deficient items. Therefore, teachers can improve the items over time, making it possible to obtain high quality tests with less effort (Muiznieks & Dennis, 1979). Johnson *et al.*, (1981), claim the goal of a computerized test item bank is to:

> . . . reduce the time an instructor must spend on the mechanical processes of writing, reproducing, and grading exams. Properly designed and operating CATC systems allow more time for two of the instructor's most important functions: selecting exam content and interpreting student performance. The computer can assume the duties of word processor, grader, and statistician. (p. 177)

As a result of using a computerized test item bank, the process of student evaluation should improve. More importantly, Hambleton, Anderson, and Murray (1983) point out that the use of computerized items banks in creating classroom tests can potentially overcome the problem of poor teacher training in item writing by making quality test items more readily available.

Another benefit can become apparent when the item bank is extremely large; security access to items can be removed. This can allow students access to items as a study aid. It also allows test items to be examined publicly and thus the community of users can judge the validity of the items (Zacharias, 1979).

Ironically, little, if any research, has been done on the effect of use of computers in test development outside of the adaptive testing area (Hsu & Sadock, 1985). Early studies at Indiana University (Moore, Prosser & Donovan, 1971) and University of California (Johnson et al, 1981; Seely & Willis, 1976) demonstrate that using a computerized test item bank to develop tests can have a significant positive influence on student performance. Recently, there have been descriptions of how test banks can help in the classroom environment (Bradbard, 1986; Peterson & Meister, 1983; Vogel, 1985). For example, Bradbard (1986) describes a simple program written in BASIC which can assist the classroom teacher in generating tests.

7. Application of Computerized Item Banks

Insight can be gained by analyzing individual cases of how the computer can be used to assist in test construction. According to Hsu and Sadock (1985), "Item banking systems tend to fall into two categories: those equipped with and designed for specific set of items, and those that require users to enter and store their own items." (p. 21). These two ideas were kept in mind for the following review of item banks. This review emphasizes computerized item banks reported since 1973 and lists them in chronological order.

MEDSIRCH

Hazlett (1973) developed MEDSIRCH, which is a computerized item bank for the storage of multiple choice items. The program was written in FORTRAN and is still operational today (Personal interview, Sept. 1986). Punch cards were used to enter text and data for each question. Fifty-seven variables were used to describe each item. Examples of these variables are: taxonomic level, degree of importance, educational level, difficulty level, biserial coefficient, answer, and number of times the item was used previously. There is even a category which asks if related audio visual material is needed. MEDSIRCH allows for the retrieval of an item, based on the student's prior knowledge about the question or no prior knowledge about the question.

Classroom Teacher Support System (CTSS)

Toggenburger (1973) described the Classroom Teacher Support System (CTSS) which was developed as a joint project between Los Angeles Unified School District and IBM. The details of the CTSS have been outlined by Byrne (1976) and Romaniuk

(1980). CTSS was initially conceived in 1969, and was based on the belief that good quality questions are key instructional tools for any teacher. Initially a bank of test items was developed to support high school history teachers, which started out with approximately 8,000 multiple choice questions. The questions were classified for retrieval based on the following attributes: course, category, difficulty level, behavioral level, and keywords. Questions are randomly selected if the number of items generated by the computer, for the attributes specified, is greater than required. Each item contains a corresponding file of statistical information with such things as difficulty level and discrimination index (i.e., the frequency with which the students chose each alternative in the multiple choice question). Furthermore, the system allows for the automatic scoring of items to take place, by means of an optical scoring device. Teachers order tests by filling out a request form and sending it to the computer center via the school district's internal mail system. The turn around time is between 24 and 48 hours. "Teachers rate CTSS as extremely helpful. They state that it saves them time and provides quality materials" (Toggenburger, 1973, p. 43).

Computerized Item Banking System (CIBS)

The Computerized Item Banking System (CIBS) was design in Britain by that country's largest test development agency (the Test Development Research Unit). The underlying premise of CIBS is that a "multiple choice test item may only be cost effective in the widest sense if it can be used on more than one occasion" (Newbould & Massey, 1977, p. 115). CIBS stores only information about the question, not the actual question. Question are stored on a manual file card system. Thus, CIBS can be regarded as a mixture of automation and systematic manual storage. New items are all pretested before being included in the bank and items must meet specific statistical criteria. That is, items are stored on the basis of item characteristics. A data file is created for each item and contains such things as:

1) the author, the year it was written, and the author's own serial number for the item.
2) which subject pretest it was in and what item number it carried
3) the syllabus topic tested
4) the skills tested
5) the type of item
6) is the item part of a group
7) most recent uses of the item
8) item statistics
9) the identity of the key

Test construction involves examining whether the item meets appropriate specifications. When the bank is large enough, a pool of items is identified and the items are examined in more detail. An optical mark reader is used to score the students' responses. In this way data for each question is easy to manage and records are easy to update. The procedure also allows for easy reference as to how teachers might use the bank.

SOCRATES

The SOCRATES system (Seely & Willis, 1976; Johnson et al., 1981) at the California State University and Colleges, has been used since 1974. It is a state-wide network system which provides access to 11 different item banks that range from a few hundred items, in some areas, to 10 000 items in first year chemistry, and 15 000 items in American History. As of 1981, there were 80 000 test questions in the SOCRATES system from 26 different subject areas. Tests are printed on local high speed printers (if available). Tests can also be ordered by phoning the central office in Los Angeles and the tests are delivered by the school district's courier service. Items can be retrieved by subject category, topic code, difficulty level, cognitive level (as defined by Bloom's taxonomy), and keywords (up to three keywords). The system also allows for linked items. Up to 10 versions of the same test with the same items randomly ordered can be obtained.

Computerized Question Banking System (CQBS)

Johnson and Maher (1982) developed a Computerized Question Banking System at the University of Leeds (England), to monitor the Department of Education's science program. The computer language they used to designed the bank was CODIL. Before any questions were entered into the bank, they underwent extensive review by instructors, pretesting, and validation. Each question has a series of label values attached to it which describe its content, such as:

LABEL	Range of Values
Category	1 to 5
Sub-Category	0 to 4
Descriptor	0-9 - defines a limited range of question types
Concept Region	A to F (Biology to Physics)
Time	time need to complete questions
Year	year in which question added to bank
Mean score	mean score from trial
Correlation	correlation with parent sub-category
Surveys	the survey years in which the question was used
KEYWORDS	
SYNOPSIS	

The initial intention of this item bank was to store the complete question on a computerized data storage facility. Problems resulted when questions with diagrams or photographs could not be stored without extensive cost. Thus a manual file system was used for such items. Any user is allowed to browse through the bank, on-line. Items are browsed based on specific labeled values. Basically, test construction works the same as browsing, where items are withdrawn from the bank based on specific label values. A pool of relevant items is obtained by this test construction procedure. A specified number of items are then randomly selected from this pool and a test is generated according to a set of specifications.

Wisconsin Item Bank

The Wisconsin State Board of Education developed an item bank from which local districts could develop tests matched to local objectives (Burke, Kaufman & Webb, 1985). The Wisconsin item bank contains 10 000 items for grades 3-12 in reading, mathematics, and language arts. These items are stored in a computerized system that has been in operation since 1984. During the first year of operation, tests were provided for 35 of the 140 available school districts. The five design specifications important in the development of the bank were:

1) item entry
2) item selection
3) test construction
4) record keeping
5) scoring/reporting

According to the designers of the Wisconsin item bank, these specifications were subsumed under three variables which can be used to classify any item bank already in existence or those in the process of being developed. These variables are:

1. the organization and structure of item collection.
2. the extent to which procedures for accessing, reviewing and retrieving items is systematized.

3. the extent of the automation of procedures

As a result, an item bank can be classified within a three-dimensional space (Figure 1). According to the interpretation by Burke, Kaufman, and Webb (1985), the Wisconsin Item Bank is located in the middle of the three extremes.

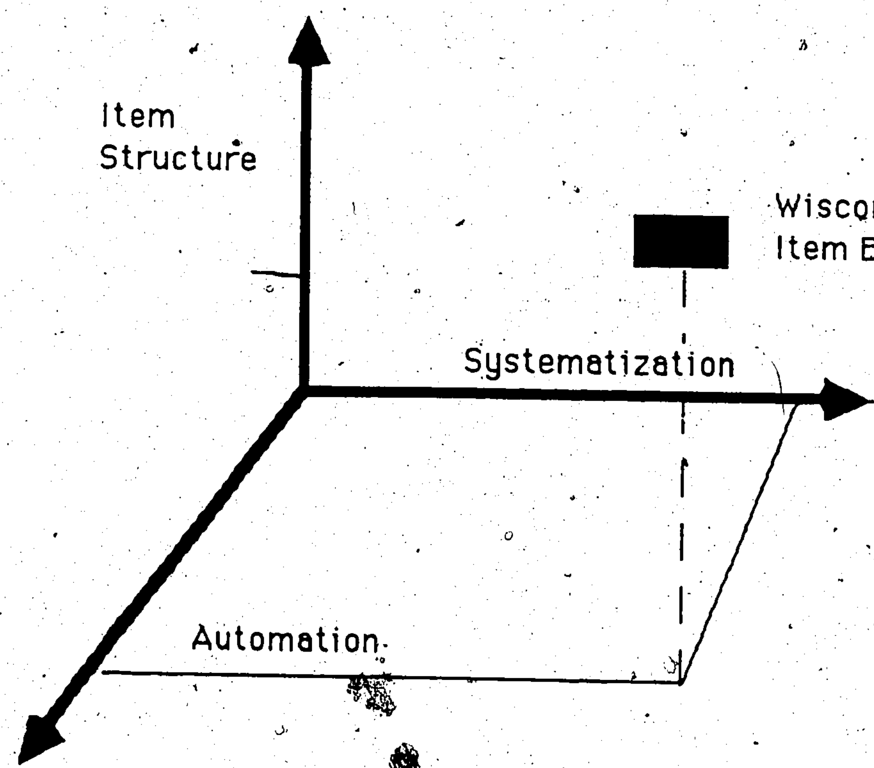

Figure 1. Item Bank Space (adapted from Wisconsin Item Bank Case Study)

For example, in the dimension of item structure, items have been reviewed to ensure that they conform to the State's prescribed standards, but many have not undergone field testing and thus cannot be classified according to statistical information. On the dimension of systemization, users are forced to follow some specific steps when working with the bank, but do have some freedom when making requests for items. Initially the automation dimension was designed to be on two computer systems. One system would handle text and graphics and the second would handle data management and statistics. Only the text and graphic section of the computer item banking system has been developed. Thus the automation is only partially complete.

CBTS

A CML (Computer Managed Learning) system, called CBTS, developed by Computer Based Training Systems Ltd. in Calgary Alberta, is widely used by postsecondary institutions in Alberta. It is run on Digital Equipment Corporation's family of VAX computers. CBTS offers a comprehensive computer managed learning environment with such features as curriculum and course organization, maintaining student records, test generation (item banking), communications, instructor managed system, and integration with a computer based instructional system. Testing can be done on-line or off-line.

Individual test banks are created for a course. Each test bank holds up to 99 modules (units). Each module can hold up to nine objectives and a maximum of 99 questions can be stored for each objective. The system can store the following question types: multiple choice, short answer, true/false, essay, and matching. The instructor can enter an algorithm which generates parallel versions of math questions. It has complete test scoring capabilities and a reporting facility. The system has the ability to include simple line diagrams in each question. Recently CBTS was upgraded to allow special character sets to be included. Test items can be grouped according to the intended type of test (quiz, unit test, final exam). An unassigned field is available so that users can select items based on a characteristic that is not defined, but could be useful, such as a cross reference to a textbook. Unfortunately, this field is only one character in length. Selection attributes are module, objective, question type, intended type of test, cognitive level, and degree of difficulty. Specified weights can also be assigned to individual items.

Identification numbers are automatically assigned as items are entered into the bank. Special instructions can be specified for each item. Items are limited to 100 lines and cannot be linked to one another. Items can be selected randomly and the system has the ability to keep items of the same type together. A very comprehensive statistical analysis is available, but few instructors use it for anything other than difficulty level (personal communication, June 1987). Very poor word processing capabilities slow down the entry and editing of items.

PETA

Nitko and Hsu (1984) developed a very comprehensive system of microcomputer programs for the classroom teacher. The system called Pittsburgh Educational Teaching Aids (PETA), is designed to run on Apple II microcomputers with at least 48K bytes of random access memory. The system is made up of three main components: 1) a student data-base with information storage and retrieval programs (as well as computational programs), 2) an item analysis component, and 3) an item banking test construction component. The three components are fully integrated, so that information used in one component can be used in another. The student database component can contain up to 10 different types of files, such as files concerning class variables or files holding variables related to student information. Seventeen individual item analysis programs are used to analyze information from the student database component. These programs provide information in three areas: 1) for improving classroom instruction by providing a summary of class performance and reporting unusual performances of a student on an item, 2) the quality of test items, such as item difficulty level and identification of poor distractor information, and 3) help in selecting items for constructing a test (items can be selected based on statistical attributes). The item banking component is made up of 12 comprehensive programs, such as searching and retrieving items based on specific criteria and estimating properties of a test based on items selected for test. The system is menu driven and allows for essay, short answer completion, matching, multiple choice, and true/ false items. PETA is not capable of storing graphics, special symbols, or subscripts. Individual test items are limited in length to 1 404 characters.

MICROCAT

MICROCAT, developed by Assessment Systems Corporation, is an adaptive testing item banking system designed for use with the IBM PC (with a color monitor). The system emphasizes testing for both diagnosis and mastery learning. MICROCAT encompasses the following four separate programs: 1) item and test development, 2) examination, 3) item and test assessment, and 4) test management. The four programs can work separately or in conjunction with one another. The system is menu driven, which makes it easy to use. Tests must be taken on-line and there is no option available for generating off-line tests. High resolution color graphics can be created with the aid of a mouse or with arrow keys. These graphics can be incorporated into any item. Each item is classified by a name consisting of one to six letters, followed by an identifier of one to three letters or numbers. Items can also be linked together, as is the case when graphs are referred to by more than one item. The entire system can be installed on a local area network and testing can be monitored from a single station. Special characters and fonts can be created by means of a font generator. Individual text can also be colored. Item response theory can be used for constructing tests from the item bank. The system also allows for test presentation programs to be written in a specialized adaptive testing language. Student data from each test is stored in a data file, which can be accessed by the item and test assessment program.

Although MICROCAT is a powerful microcomputer item banking and testing program, it is too restrictive for a typical school classroom. Very few schools have the necessary equipment (IBM local area networks with color monitors) to run this system. Furthermore, the system is designed for on-line adaptive testing. Few teachers have the required basic knowledge of item response theory. Hsu (1986) summarizes this point nicely, "Whoever can [have] access to microcomputers may have a chance to apply item response theory into practice. A great deal of effort is still required, however, before these applications can be transported from the hands of researchers into the hands of teachers" (p. 18).

MicroTest II

MicroTest II, developed by Chariot Software Group (1984), is a microcomputer program designed for the Apple II and Apple Macintosh computers. The Macintosh version, which was formally evaluated by this researcher, uses an interface involving pull down menus and a mouse. The developers call a file of test items a "Test Data Bank". Items in the Test Data Bank are categorized according to the following attributes: book title, author, publisher, chapter title, and question number. Individual items are stored according to the chapter to which they relate. Chapters must be created one at a time. This procedure closely corresponds to the practice of book publishers who supply test questions based on a particular book.

One floppy diskette is required for each Test Data Bank and a hard-disk drive cannot be used. When individual tests are created, they must be saved on the same disk as the MicroTest II program. Three question types are supported. They are multiple choice, true/false and free format (essay). Text editing features often found in common word processing systems are available, but MicroTest II does not take advantage of the graphics capability of the Macintosh and thus diagrams cannot be included with any item. Different fonts are supported, but features such as bold, underline, superscript and subscript are not. Only one type of font can be used for each printed test. No items statistics are stored and tests must be created by retrieving items from a specified chapter. Tests can be exported as text files to other word processing programs for more sophisticated editing. According to teachers presently using this program, this ability to export tests to a word processing program was the best feature of MicroTest II.

measureUP

Recently Logic eXtension Resources (1986) released a microcomputer based test banking system called measureUP. Like MicroTest II, measureUP runs on the Apple Macintosh computer. The program is designed to take advantage of the Macintosh user interface. It is similar to MicroTest II in that both pull down menus and a mouse are used. Test items are classified by objective. Each objective must be assigned a unique identifier of one to nine characters (letters or numbers) in length and saved within the test bank file. All items in the test bank file which relate to the same objective are further identified by a sequence number; a maximum of 999 items can be classified within each objective. The program allows test bank files to use the maximum disk storage space available (hard disk or floppy disk). Test items can be created and easily modified using the same full screen editing features normally found in word processors available for the Macintosh. MeasureUP allows graphics to be stored within either the item stem or alternatives. The program also supports special fonts, symbols and character formatting features such as bold, italic, superscript, and subscript. The following item types are available: multiple choice, true/false and free format. Minimal statistical information can be stored with each item. However, this information must be entered manually. The date of the last revision of each item is automatically saved. Items can be linked to one another. MeasureUP also provides a comprehensive database with boolean operators which can be used for searching and retrieving items from an item bank file. For example, items can be selected by objective, sequence number, text in either stem or alternative, keyword, special notes entered along with each item, statistical information, and questions which contain pictures. Tests can be easily assembled (via the data-base selection feature) and items can viewed before placement in a test. Multiple versions of a test can be printed, with the items in the same order and the alternatives scrambled. Test directions, along with footers and headers, can also be included with each version of the test. Items can be both exported to, or imported from, other sources. This feature aids in compatibility with existing computerized item banks and helps to prevent the drudgery of retyping items in order to make them into a measureUP item bank file.

Netherlands Network

Advances in microcomputer technology have allowed for the development of microcomputer networks. Baker (1984) describes the development of a nation-wide test scoring and reporting network in the Netherlands, where each school is being equipped with 1) a desk top card scanner, 2) a microcomputer with disk drives, and 3) a printer. After tests are completed by the students, the answer cards are scanned locally and the results stored on disk. Through the use of a modem, these files are later transferred to a central computer. After the tests are scored, the items undergo analysis and the results are transferred back to the school where they are printed on the school's printer. According to Baker (1984) this allows for the comparison of tests and items on a larger more comprehensive basis while, at the same time, providing teachers with flexibility to modify instruction based upon quickly obtaining results. Details provided by Baker about the system are sketchy and incomplete. There is no mention as to how and where the tests are actually constructed. For example, is the test down-loaded from the central computer in finished form, or do the teachers make up these tests from a given set of questions provided from a central bank?

8. Important Features of Computerized Test Item Banks

As described in the literature review, features of computerized test item bank systems are extremely diverse relative to their function and design. Many large mainframe computer item banks can store thousands of items (Brown, 1982). Some microcomputer systems, however, store less than 75 items (Deck, Nickel & Estes, 1985). The Portland Test Item Banking System (Hathaway, Houser & Kingsbury, 1985) and the Computerized Question Banking System (Johnson & Maher, 1982) store items by means of a manual file system, whereby the description of the items is completely computerized. The link between the two different storage facilities is established through a unique I.D. number. Some banks only allow multiple choice items (Newbould & Massey, 1977), while others try to include every possible item type, with as many variations as possible (Denney, 1973).

Computerized test banking systems should be capable of handling special symbols which are often required in test questions, such as those that occur in chemistry (Vogel, 1985). Unfortunately, most systems, whether microcomputer or mainframe, are not capable of handling special symbols or of storing graphics (Baker, 1986; Byrne, 1976; Hiscox, 1984; Johnson & Maher, 1982; Nitko & Hsu, 1984).

Some teachers only use word processing programs to create and store test items and to develop tests (Collins, Ellis, Fiske, & Genco, 1985). For example, the Educational Support Services at Eastern Maine Medical Center used a word processing program, IBM Displaywriter, so that "all current test items from the nursing courses [could be] stored on disks and print-outs were available for faculty use in construction of individual tests" (Collins, Ellis, Fiske, & Genco, 1985, p. 349). Problems can arise because the number of disks required to store test items increases quickly, thus potentially becoming difficult to manage. In the case of word processing programs, tests are often created by re-entering items or by cutting and pasting items from other files. According to Bardbard (1986, p. 261) the following are some of the disadvantages of using these methods to develop tests: 1) the procedures are often time-consuming and cumbersome, 2) duplication of effort is prevalent, 3) errors are introduced where there were none before, and 4) multiple copies of the same items are produced, therefore storage space is wasted. As a result, word processing programs often lead to more work in developing and storing test items.

Curtis (1972) points out the importance which book publishers can have in item development, since many tests are being developed around specific content covered in certain textbooks. According to Baker (1986, p 410), "the current trend is for these tests to be stored on a floppy disk accompanied by microcomputer software for the selection and printing of the appropriate test". This feature is often be overlooked when item classifications schemes are developed.

Johnson and Maher (1982) recommend an audit trail be developed in tracing the revision of items. Newbold and Massey (1977) point out that it is important to monitor test items by examining features such as: who wrote the items, where were they developed, and when were they developed. These features should help to guide administrators in improving the bank (Newbold & Massey, 1977).

Millman and Outlaw (1978) developed a computerized test item creation program where items are created through the use of previously stored algorithms. These questions are often referred to as formula generated questions (Lippey, 1974) or parallel test questions. These formula generated questions allow for repeated testing, practice exams, control of cheating, make-up tests, and tailored tests (Millman & Outlaw, 1978). Furthermore, storage space can be greatly reduced since one program can be used to develop many questions.

Wright and Bell (1984) emphasize that items should be drawn from a test bank based only on a calibration scheme. That is, "item calibrations are obtained by applying a probabilistic model for what ought to happen when students attempt an item" (p. 333). Other writers also emphasize the need for items to be included in a test based on their statistical properties (Baker, 1973; Hambleton, 1984; Hathaway, Houser & Kingsbury, 1985; Nelson, 1984). Valid and reliable items can be effective in tracking student performance (Johnson & Maher, 1982) or in establishing cutoff scores for a program which requires a specific level of achievement before certification (Legg, 1982).

A variety of question types (multiple choice, short answer, etc.) provides flexibility to an item bank (Denney, 1973). On the other hand, the smaller the number of types of questions, the greater the simplicity of system design and operation. Byrne (1976) points out that the time required to score and analyze test items will increase as the number of item types increase.

Computerized item banks usually involve the storage of items according to predetermined item attributes (Stodola, 1974). CTSS in Los Angeles uses six variables to store item characteristics (Toggenburger, 1973), while MEDSIRCH uses 57 variables (Hazlett, 1973). Attributes usually vary from "assigned attributes; attributes which are intrinsic to, or implicit in, the question itself" to "measured attributes; attributes which are dependent upon data collection from actual use of the item" (Muiznieks & Dennis, 1979, p.18). The following assigned and measured attributes were identified by Muiznieks & Dennis (1979).

Assigned attributes:

- identification number
- question
- correct answer
- wrong answer(s)
- subject matter
- estimate of item difficulty
- location in the curriculum
- keywords
- guidelines for grading the question
- comments about the question

- cognitive level
- name of item author
- question type
- references to related materials
- instructional objectives
- cross-reference to other characteristics

Measured attributes:

- difficulty index
- discrimination index (this may depend on the type of item)
- effectiveness of distractors
- number of times item was used
- last time item was used
- previous version of the item
- time required to answer the item (on-line or off- line parameter)

In general, a number of important features exist which computerized item banks should incorporate if they are to function successfully. Items should be categorized using a meaningful classification scheme. The classification should not center around quick and easily retrievable items (Hsu & Sadock, 1985), although Hambleton (1984) and Hiscox (1983) point out that this feature is important on microcomputers. Instead, proper classification should emphasize the quality of test items (Wright & Bell, 1984). This will aid in the construction of more valid and reliable tests (Hsu & Sadock, 1985). Many of the attributes identified in this section, or in the previous section, could be used in a classification scheme. Many item banks use classification schemes based on objectives or topic (units), and statistical properties. Hiscox and Brzezinski (1980) point out other classification schemes. These schemes often center on the subject matter of the item itself. In essence, the classification scheme will focus on the purpose for the bank and the local setting (Estes & Arter, 1984).

An item banking system should have the ability to analyze item and test data. This requires the necessary equipment, such as an optical card reader, to carry out test scoring and reporting. Researchers in the early 1970s realized that if this feature was not automated and easy to use, the teacher would not calculate item analyses by hand (Denney, 1973). A number of microcomputer systems are capable of this procedure

(Nelson, 1984; Deck, Nickel & Estes, 1985). Due to the variety of item data which can be obtained, the system should be flexible enough to adapt to teacher needs (Hsu & Sadock, 1985). Unfortunately, this is still a difficult goal to accomplish on most microcomputer systems (Nitko & Hsu, 1984).

The item bank should be easy to maintain and update. This means items should be easy to create, modify, store, retrieve, and delete. The powerful editing features often associated with word processing systems have shown to be very beneficial when editing items within a bank (Collins *et al.*, 1985). As Hsu (1986, p. 17) has recently noted, " a good item banking system should be equipped with such [word processing] capability". Furthermore, computerized item banks should permit the storage of items with associated diagrams (Burke, Kaufman, & Webb, 1985). There have been a number of attempts to deal with the graphic problems on mainframe systems (Denney, 1973; Stodola, 1974). The Wisconsin item bank, which uses an expensive Xerox Star system, successfully uses graphics within item text (Burke, Kaufman, & Webb, 1985). Early microcomputers suffered from limited technology and, as a result, programs designed for those machines made little or no use of graphics. Hiscox (1983) points out these early programs often increased the workload involved with creating graphics for a test. MircoTIM (Vogel, 1985) has special character sets available to graphically represent chemical structures, but this feature is still restrictive. A number of recent microcomputer programs have successfully included graphics with test items; this includes MICROCAT which runs on the IBM PC, and measureUP which runs on the Macintosh.

Another feature of a computerized item bank has to do with the generation of tests. Here it is important to keep in mind Weber's (1984) point, that the development of an item bank must be done in conjunction with the test development process, which it must support. Tests should be generated in conjunction with the attributes which are associated with each item. That is, items will be retrieved from the test bank based on specific criteria or upon a logical relationship between two or more criteria (Hambleton, 1984). For example, a person could first search the test bank, based on specific general criteria (e.g., objective), and then further narrow the search by selecting items based on statistical criteria (e.g., item difficulty). Test generating features often include the ability to review an item before it is placed in a test, allowing the user to specify the date of the test and specific instructions for the students, production of answer keys, the ability to order and/or reorder items before the test is printed, and the ability to adjust the page lay-out.

A final feature which seems to be overlooked in the literature is the ability of the system to accept (import) items from other computer systems and programs. Recently, word processing systems have been used to create and store items (Collins, Ellis, Fiske, & Genco, 1985). A computerized item bank which allows items to be imported, without having to retype the items, could save users a tremendous amount of time. This procedure would usually involve the use of a data transfer system, such as phone modem or local area network. There are a large number of existing item banks from which quality items can be obtained (Brzezinski, Powers, Tamura, & Hiscox, 1980; Dawson, 1987). For example, textbook publishers are beginning to supply tests and test items stored on floppy disk instead of including the materials in an instructor's manual (Baker, 1986). Some programs can be purchased with corresponding item banks already developed (Deck, Nickel & Estes, 1985). As Hiscox (1983) states, reuse of these external items can have economic benefit because of the saving on item development costs.

Presently, few microcomputer software packages exist (if any) which incorporate all of the above features (Hambleton, 1984). Programs have been designed to do many of these features separately, but the problem of interfacing these programs is often very difficult (Baker, 1986).

9. Considerations in *Designing* a Computerized Test Item Bank

The preceding review of the literature has summarized the purpose for testing, the development of quality exams, how a computerized item bank might help teachers, various applications of existing computer systems, and features which are generally present in a variety of computerized item banks. These are all important aspects that need to be considered when selecting a computerized test item banking system that meets the needs of the classroom teacher.

Baker (1986) outlined two approaches which may be used when employing a computerized test item bank in the school environment. The first is a decentralized approach, in which microcomputers act as stand alone item banking systems. This method has two advantages in that it allows the classroom teacher easy access for test construction, and teachers can develop a personal feeling of ownership. The disadvantage of employing a decentralized item banking system is the potential for decreased quality control over the items that will be used. Item banks created by individual teachers, or individual departments within the school, may contain "varying levels of quality and uncertain commonalties across the curriculum" (Baker, 1986, p. 408).

The second approach to implementing an item banking system is that of centralization. Each microcomputer can act as terminal, which can be used to download pools of items from the large centrally located mainframe. In this way, tests can be created locally but quality control of items can be maintained at the centralized location (Baker, 1986). Nevertheless, teachers can likely manage an item pool tailored to their needs far better than any central office bureaucracy, especially since teachers make daily decisions in response to local objectives, changing school environment, and the different ability levels of their students. Therefore, a centralized approach to computerized item bank may not be the most appropriate. On the other hand, for a decentralized computerized item bank to work, teachers need the necessary tools (appropriate microcomputers and software) to successfully construct tests.

Arter and Estes (1985), Estes and Arter (1984), and Millman and Arter (1984) have tried to develop guidelines for the design of a item banking system. According to Millman and Arter (1984, pp. 323-327) design considerations related to the development of a test item banking system can be broken down into a number of specific categories. There are a number of important questions associated with each of these categories.

1) Questions related directly to the *items* themselves. Here the focus is on the extent to which the items are acquired or developed. Such concerns are, should the teachers develop their own items or use items developed by others? Which types of items are permitted? How large will the item collection be? How will the items be reviewed before placement into the bank?

   Other questions relate directly to the classification scheme used for the items. Concerns in this area center on, the way subject matter should be classified, what information about the item should be stored in the bank, and how the bank is managed. Is the capacity large enough for updating items and classification schemes? Can diagrams or special characters be stored with the item or must they be stored separately? Can items be linked to other items for easy recall?

2) The second category concerns how *tests* will be assembled, administered, scored, reported, and evaluated. What parameters exist to control item selection? How much control will the user have in formatting and printing the test? Can scoring keys be generated? Will statistical data for both the test and the items be calculated?

3) The third consideration involves which *item banking system* to use. Will the system be commercially acquired or will it be privately developed? What

software and hardware features are required (e.g., if a microcomputer is chosen, what limits will there be on type of text, size of the bank, graphics, and test development options)? Will users require special training? Who will have access to the system and what level of security is necessary?

4) The fourth consideration involves *determining the way in which the system will be used.* Is a function of the bank to improve instruction? Is the bank intended to store items? Is the bank to be used in certification examinations? Will the bank be used to define curriculum performance?

These questions can help determine guidelines from which a microcomputer test item banking system could be built.

If computerized test item banking programs are to be used as tools for teachers to improve classroom instruction, develop quality tests, report student test results, and evaluate student learning, then serious consideration should be given to what classroom teachers need. The criteria for designing microcomputing testing software for classroom teachers must center on the needs of these teachers.

As a result Hsu and Nitko (1984) point out that:

> . . . if tests and other assessments are to be used to inform a teacher's classroom instructional decisions, they need to be linked closely to the daily practise of the teacher, rather than to the broader, developmental goals by which the entire educational enterprise is to be judged. . . Computer-assisted testing programs, whether they are microcomputerized or not, must maintain this ability to accept teacher's style of testing if they are to impact on daily instructional practices. . . . [N]ew technological tools must be teacher-oriented and, at the same time, assist the teacher to focus his or her instructional efforts on students. Although the technology may be teacher-oriented in its mechanical operations, the outcome of using technology should be to increase a teacher's capacity to be public-centered. A microcomputer innovation must be designed, therefore, to facilitate a teacher's role as a teacher. One consequence of this orientation is that computer technology must be designed to assist a teacher in making decisions and must not make decisions for the teacher. . . . It views the teacher as the responsible instructional agent. Because the teacher is in charge, the technology should be designed to adapt to the teacher's individuality, rather than requiring the teacher to adapt to the technology. One implication for computer software design that follows from the

> criterion of adaptability to teachers needs is that computer software must be capable of being used by virtually any teacher. . . . [T]his teacher-oriented, pupil-centered, user-friendly point of view about technology. . . [suggest] ways to design computer-assisted classroom testing system. (p. 16)

The goal is to use existing technology to "facilitate a teacher's role as a teacher" (Nitko & Hsu, 1984, p. 378). Knowing the design considerations involved in the development of a microcomputer test item bank can help to successfully implement it into the classroom. For example, in the initial design phase of Hsu and Nitko's (1984) PETA program, classroom teachers were asked what features they thought should be included in the system.

Any implementation of the microcomputer in the classroom environment, should diminish the teachers' workload rather than increase it (Baker, 1986). Early researchers, such as Denney (1973), noted that any level of innovation must allow teachers to begin at their own professional level and grow as their skill improves. There must be an evolutionary process among the teaching staff, with some using only the printed tests produced by the program, and others becoming involved with the daily operations of the system.

10. Considerations in *Choosing* a Computerized Item Bank

What sort of computerized test item bank programs are presently available for use in the classroom environment and do these programs meet teachers' needs? The previous literature review focused on both mainframe and microcomputer test item banks. Educational researchers such as Hsu and Nitko (1983), Ju (1984), and Deck, Nickel and Estes, (1985) developed instruments to assist in the formal evaluation of existing item banks. For example, Hsu and Nitko (1983) surveyed 56 microcomputer software packages which assisted in testing and found only 18 packages capable of item banking.

Of those 18 packages, 11 were capable of printing tests and, of these, only one had an item analysis feature. Hsu and Sadock (1985) refer to a survey carried out by Ju (1984) in which a "user's evaluation form" was developed to assist teachers in judging microcomputer testing aids. This form is somewhat limited because it examines general attitudes of teachers rather than the particular features of the test item banking system.

In a comprehensive study, Deck, Nickel and Estes (1985) formally evaluated 18 microcomputer test item banking software programs (a recent literature search, done at Northwest Regional Educational Laboratory indicated very few additional programs [personal communication, June, 1987]). These programs were meant for either the Apple II or IBM PC, and cost anywhere from $20 to $2400 US. To assist in this evaluation, Deck, Nickel and Estes developed an instrument which outlined the major features usually associated with computerized item banks. Features of these microcomputer item banks were grouped according to six categories: 1) item bank description: general features the item bank might have, 2) item bank maintenance: how easy is it to edit and maintain the item bank, 3) test assembly: what features are present to aid in the construction of a test, 4) administration and scoring: are features available for analyzing tests and items, 5) student record keeping, and 6) ease of use. Table 1 has been developed to summarize the results. These results show that there is a great deal of variation in existing microcomputer test item banks.

Table 1

Summary of Results Obtained from Deck, Nickel and Estes (1985)

Item Bank Description		*Item Bank Maintenance*	
Features	**# out of 18**	**Features**	**# out of 18**
May purchase item banks	5	Create and edit items	17
Use various response formats	9	Use full-screen editing	5
Link text passage to items	5	Store usage history	8
Store classification scheme	8	Calculate item statistics	3
More than 4 lines in item stem	15		

Table 1 (continued)

Test Assembly		*Administration and Scoring*	
Features	**# out of 18**	**Features**	**# out of 18**
Select by item number	13	Administer test on-line	9
Select randomly	10	Support scanner	0
Select by classification	8	Computer totals on test	8
Print multiple forms	7	Determine objective mastery	2
Support special characters	4		

Student Recordkeeping		*Use Factors*	
Features	**# out of 18**	**Features**	**# out of 18**
Use objectives approach	0	Comprehensive manual	11
Use gradebook approach	5	Easy to use	12
		Reasonable performance	11

More recently Broduer (1986) formally evaluated features of seven microcomputer test generating programs. Broduer's features were categorized along the following lines: 1) hardware requirements, 2) item pool size, 3) item construction features, and 4) test formatting features. As was the case in the other evaluations of microcomputer test item banks, Broduer also found variation among the different programs. As a result of the variation in the features found in available item banking programs, it is difficult to select one which best fit the needs of the "average classroom teacher".

Current development of computerized test item banking systems is largely being done on microcomputers (Hsu & Sadock, 1985). Furthermore, it appears that little or no research is ever done before a computerized test item bank is designed, developed, and marketed (Deck & Estes, 1984). According to Hsu and Sadock (1985), there are three reasons why this has occurred: 1) timing is very important in a competitive market, thus delays resulting from market surveys will increase the cost, 2) technology is changing at such a rapid rate, the system must be marketed quickly or it will become obsolete before it is marketed, and 3) users are often so fascinated by the technology that they fail to realize the implications it has in education.

Estes and Arter (1984), in developing guidelines for state and local computerized item banks, concluded that the utility of an item bank can be "maximized if the designer knows in advance what features users want, what features actually will be used, who to consult in [illegible] design phase, how to establish credibility, what turn–around time is needed, and [illegible]rces are necessary and available to develop tests using items from the item bank" ([illegible]5). To date, the only preliminary study to assess which features are important in microcomputer test banks was carried out by Pine (1986) at Arizona State University.

Pine's study focused on testing at the university level in the faculties of Business, Psychology, and Nursing. Twenty-seven features of microcomputer test item bank programs were evaluated by professors (total N = 48) to determine which would be most useful in the current testing situations. These features were grouped into five categories: test preparation formats, test item types, test assembly methods, forms and limitations of test generation, and other information stored with test items. Results show that 97.9% of all respondents use multiple choice items, 50% use true/false, and 95.8% want printed tests rather than on-line delivery on computer. A majority (75%) want to review the text of each item before placement on the test. A difference in subject areas occurred in the assembly/selection method with 60% of the Business, 100% of Psychology, and 22.2% of Nursing indicating they select items by chapter. Selection of items based on objectives was found to be important in Nursing (72.2%), but not so in Business (8%) and Psychology (0%). Other differences between subject areas occurred in selection of items based on content, with the importance level varying from 88.9% in Nursing, to 36% in Business, and 20% in Psychology. Most respondents (60.4%) developed tests that ranged between 50-100 items. Finally, although 81.3% of all faculty members felt that statistical analysis is important, only 50% of the faculty members selected test items based on item difficulty level.

CHAPTER III

RESEARCH DESIGN

This chapter describes how the present study was conducted. The first section describes the development of the instrument, which was devised to measure the study questions. In section two the sample and setting are addressed, and the procedure for data collection is described. Finally, data preparation and the methods used for statistical analysis are presented.

1. Development of the Instrument

Borg and Gall (1983) point out that survey research can provide data to help "predict educational needs" (p. 405). Furthermore, use of survey research by a questionnaire provides a feasible method by which to obtain the views of a population in a selected geographic location. For this study, a four-page survey questionnaire and a one page cover letter were developed (see Appendix B for copy of survey). The questionnaire is divided into two sections. The first section is concerned with general information, while the second section focuses on rating the level of importance of individual item bank features. In the first section, specific responses required either a written statement or a checkmark from the respondents. In the second section of the questionnaire, a four-point rating scale was adapted from the *Survey Research Method* (Mason, *et al.*, 1983) and included the following categories:

- Very Important
- Important
- Somewhat Important

- Not Important
- Don't Know

The "Don't Know" category was an addition to the scale and was included for those teachers who did not understand the accompanying statement or did not have the experience on which to base an opinion. Additional space was provided at the end of the questionnaire for any comments individual teachers might wish to make. None of the questionnaires were coded with any special marking to ensure that anonymity of responses was maintained.

Development of the questionnaire was guided by seven factors.

1) Relevant information was obtained by reviewing the literature regarding various features presently available in most computerized test item banks. The literature review focused on computerized test item banks which run on either mainframe computers or microcomputers. This helped determine what features might be important in a computer–based test item banking system.

2) A review was conducted on a number of locally available computerized test item banks (i.e., CBTS, measureUP, MICROCAT, and MICROTEST II). This allowed this researcher to have direct experience with each system and provided some insights into the problems that teachers might face in developing tests with these computerized test item banks.

3) Information was obtained from classroom teachers who presently use computerized or manual test banks. This information was obtained by personal communication with teachers and helped the researcher in two

respects: a) the unique approaches each teacher utilizes in organizing tests and test questions were outlined, and b) the questions used in the questionnaire were refined so that they would be concise and easy to understand.

4) The following format recommendations for questionnaire development were made by Borg and Gall (1983);

   a) Make the questionnaire attractive; use colored paper
   b) Organize and lay out question so the survey is easy to read.
   c) Number the questionnaire items and pages
   d) Brief, clear instructions, in bold print
   e) Organize questionnaire into a logical sequence, so related items are grouped together
   f) Begin with a few non–threatening items
   g) Short items are preferable and technical jargon should be avoided
   h) Avoid biased or misleading questions

Although these are simple and straightforward guidelines, they can greatly influence the return rate of the questionnaire (Borg & Gall, 1983).

5) The fifth factor for the development of the survey questionnaire was the theoretical basis on which the feature categories were defined. The features were compiled and placed into the three categories: *item structure*, *systemization*, and *automation*. These three categories are based on the theoretical dimension used in the development of the Wisconsin computerized test item banking system. Some features overlapped from one dimension into

another dimension, but were placed in only one dimension. For example, although course, unit, subject, objective, and cognitive level could have been placed in the item structure dimension, they were placed on the systemization dimension to avoid repetition and to keep the length of the survey within reason. The following features were selected as important for each of the three dimensions:

a) Item Structure (See Figure 2).

- Author or source of item
- Diagram can be included as part of the stem
- Diagrams can be included as a part of the alternatives
- Availability of special character sets (symbols, e.g. Σ)
- Availability of character enhancements (e.g. bold, italic, underline, outline, superscript, subscript)
- Individual teacher comments about an item
- Cross reference of items to specific pages in text books
- Cross reference of items to audio visual materials
- Assignment of teacher specified weights to items (how many marks is the item worth)
- Average completion time per item
- Items grouped according to the intended type of test (e.g. quiz, unit exam, final)
- A series of items linked or associated with one specific item, (e.g. three questions relating to a diagram)
- Parallel (similar) items generated by computer

Figure 2. Features Categorized According to the Dimension: *Item Structure*

b) Systemization dimension was broken into the following two parts: i) features related to selecting items from the test item bank either to review the items or to create a test [See Figure 3], and ii) general features available after selection of test items from the bank has been made [See Figure 4].

- Select by topic (Unit)
- Select by instructional objective
- Select by cognitive level
- Select by item type (e.g. multiple choice, short answer, etc.)
- Select by difficulty level
- Select by discrimination index
- Select by unique identification number (I.D. number)
- Select by visual inspection of items on computer screen
- Select items randomly
- Select by keywords in stem
- Select by keywords in alternative responses
- Select by diagram used in stem
- Select by diagram used in alternative responses
- Select by reference to book or book chapter
- Select all items in the entire bank for printing

Figure 3. Features Categorized According to the Dimension: *Systemization part I*

c) Features in the Automation dimension deal with those features that are automatically maintained by the computer (see Figure 5).

6) The sixth factor was validation of the survey instrument by a panel of experts. Early versions of the questionnaire were developed and reviewed by a number of professors within the Faculty of Education at the University of Alberta and an Educational Psychology PhD graduate student specializing in testing and measurement. The questionnaire was then revised on the basis of the responses and comments received. Once revised, the questionnaire was pre-tested with six urban area high school teachers, from six academic subject areas (English, Social Studies, Biology, Chemistry, Mathematics and

Physics). Questions which this group of teachers felt were ambiguous were then revised to eliminate this ambiguity.

- Easy editing of item text
- Easy editing of item graphic
- Select a previously used test.
- Display a copy of the test before printing
- Provision for special directions to be included with different test sections
- Generate a set of tests, where different items are used in each of the tests generated but based on same criteria
- Generate a set of tests where the same items are used but the alternatives are scrambled
- Generate a set of tests where the same items are used but the order of items is scrambled
- Keep together items of the same type (e.g. all multiple choice items together)
- Keep together items on the same topic
- Ability to reorder items before printing
- Ability to have special headers & footers printed
- Ability to modify page setup, line spacing and margins
- A variety of formats for numbering questions
- Print answer keys for all versions of the test
- Ability to merge together different test item banks

Figure 4. Features Categorized According to the Dimension: *Systemization part II*

7) The final factor was a carefully designed covering letter that accompanied the survey instrument. This letter briefly stated the purpose of the study. It also indicated that the teachers' responses were important in determining which microcomputer item banking system would best suit their needs. A definition of a *test bank* was provided. As an incentive, the letter informed the teachers that a free demonstration of a microcomputer test item banking system would be available to them at a later date.

- A unique identification number should be automatically assigned to each item when it is placed in the item bank
- A record of previous versions of each item should be maintained by the computer (modification history)
- Date of last revision for each item
- Number of times each item has been used in other tests
- Last date item was used on a test
- Security access by password
- Difficulty level
- Discrimination index
- Distractor information
- Means for high, median and low group
- Number of times response to each item was omitted

Figure 5. Features categorized according to the dimension: *Automation*

2. Description of the Sample

The intent of this study was to focus on high school teachers from the academic areas of English, Social Studies, Biology, Chemistry, Mathematics, and Physics. The reason for focusing on teachers of academic rather than non-academic subjects, at the high school level, was twofold: first, tests are often longer and more difficult to make up in these subject areas because a great deal of course content is covered (Alberta curriculum guide); and, second, the 1983 re-implementation of the Alberta Grade 12 Diploma exams has placed a great deal of stress on teachers from these subject areas. Teachers are being judged on the basis of their students' scores (Kolmes, 1983). Therefore, this researcher perceived that a computerized test item bank was needed more immediately in these subject areas.

The subjects for this study were high school classroom teachers from four urban Alberta school districts. Only one high school was available for surveying in each of two school districts. These school districts will henceforth be referred to as *school district one* and *school district two* respectively. There were four high schools available for surveying in the third school district henceforth to be referred to as *school district three*. The fourth school district, henceforth to be referred to as *school district four*, had seven high schools participating in the survey (see Table 2).

Prior to distribution of the questionnaires, the exact number of teachers in each of these high schools who taught in the academic subject areas was not available to the researcher. To ensure an adequate sample size provision was made for teachers in subject areas such as Vocational Education and Business Education, to be involved in the study. The survey questionnaires were distributed and collected in the following manner:

1) School districts one and two were contacted by phone to obtain an estimate of the number of teachers available for the survey. An appropriate number of survey questionnaires was then forwarded to the schools by courier. The principals at each of these schools then placed the questionnaires into the participating teachers' mailboxes. The cover letter attached to each survey questionnaire specified the desired completion date of the survey, along with a request that the questionnaire be returned to the school's general office. The researcher then contacted both high schools by phone and the completed surveys were returned to the researcher by courier.

2) The method used to survey the four high schools in school district three were similar to those specified for school districts one and two. The only exception was that the researcher personally dropped off and picked up the surveys at the participating schools.

3) The survey of seven high schools in school districts four was carried out with the support of the computer consultant and the supervisor of curriculum studies in that school. In this case, surveys were distributed to each school by the school district's computer consultant via the school district's internal mail system. In addition to the covering letter attached to each questionnaire, a supplementary letter was also sent to each high school by the computer consultant informing the principal of the study. This supplementary letter specified how the questionnaires should be distributed (see letter in Appendix A). Principals of participating schools in this district were phoned by the computer consultant to remind them to collect and return the completed survey questionnaires.

Table 2 outlines the response rate for the participating school districts.

Table 2

Questionnaire Response Rate

School District	Number of Schools	Teachers Available	Questionnaires Completed	Response Rate%
1	1	25	12	48.00
2	1	50	47	94.00
3	4	60	39	65.00
4	7	215	101	46.97
Total	**13**	**350**	**199**	**56.85**

3. Data Preparation

In the general information section of the questionnaire (Appendix B), questions five through eight were coded with numeric equivalents so that inferential statistical techniques could be used with data. All responses to these questions were assigned integer values in ascending order beginning with one and ending with the number of the final response. The integers were assigned in ascending order either from top to bottom, or left to right, depending on the circumstance. Question nine was assigned values in descending order rather than ascending order.

- Questions 11 to 65 which dealt with individual item bank features, were scored so that a response of "Very Important" was assigned a score of four, "Important" a score of three, "Somewhat Important" a score of two, and "Not Important" assigned a score of one. The "Don't Know" response was treated as missing data and, therefore, had no numeric equivalent assigned.

4. Statistical Procedures and Analysis

Kerlinger points out that frequency analysis is a valuable method of expressing results obtained from survey research (Kerlinger, 1979, pp. 151-157). Therefore percentage scores are reported for questions in the general information section of the questionnaire. Means and standard deviations are used to describe the responses given for those questions concerning features of a computerized test item bank. The percentage of individuals indicating "Don't Know" is reported in a graph. A t-test was used in cases involving two group means, and an analysis of variance was used to test for significant group mean differences in cases when there were more than two groups. Post-hoc Scheffe tests were applied to provide specific comparisons between groups, to determine which groups differed significantly from each other.

According to Kerlinger (1979), "factor analysis is an analytic method for determining the number and nature of the variables that underlie larger numbers of variables or measures. It tells the researcher, in effect, what tests or measures belong together—which ones virtually measure the same thing, in other words, and how much they do" (p. 180). In this study, factor analysis was used to find groupings of the features found in a number of computerized test item banks.

CHAPTER IV

ANALYSIS AND PRESENTATION OF RESULTS

The presentation and analysis of the results has been divided into five sections. The first section provides a descriptive view of the characteristics of respondents. The second section focuses on the ratings respondents gave to features of a computerized test item bank; section two is divided in two parts: the features which teachers found important, and the proportion of teachers who responded "Don't Know" to these features. Section three makes comparisons between subject areas, to determine if significant differences are present. Also in this section three, specific questions which were identified as being important in this study are analyzed. The fourth section describes a number of the open ended comments made by the teachers. Section five employs factor analysis to establish factorial validity of the questionnaire items.

1. Characteristics of Respondents

Out of 350 questionnaires distributed, 199 were completed, and returned. This represents a response rate of 56.86%. Table 3 outlines the frequency and percentage of questionnaires received from each high school subject area. Proportionally, teachers from the Mathematics subject area returned the highest percentage of questionnaires (24.12%); the lowest was from teachers in the Second Language subject area (1.51%); one questionnaire was returned with no subject area specified.

Table 3

Questionnaire Return Rate by Subject Area

Subject Area	*No. of Teachers*	*Percentage*
English	33	16.58
Social Studies	39	19.60
Biology	16	8.04
Chemistry	17	8.54
Math	48	24.12
Physics	12	6.03
Second Languages	3	1.51
Business	8	4.02
Vocational Education	22	11.06
Unknown	1	.50
Total	**199**	**100.00**

One questionnaire from the English subject area was dropped from the survey because the respondent did not adhere to questionnaire guidelines.

The 199 teachers in this study who completed the questionnaires had taught an average of 15.94 years. Mathemathics teachers had the highest average number of years of teaching (18.79). The number of university courses taken, which focused on test construction, varied as follows: 38.78% of the teachers had no such courses, 41.84% had one course, 14.29% had two courses, and 5.10% had three or more courses. The types of test items which teachers said they used on their tests varied according to the subject area in which they taught. Table 4 shows the percentage of different test items used in various subject areas. Overall, teachers tended to use multiple choice (44.45%) and short answer (33.29%) test items the most.

Table 4

Percentage of Test Item Types used in Different Subject Areas

Subject	Essay	Fill in the Blank	Multiple Choice	Matching	Short answer	True False	Other
English	47.34	4.91	21.63	5.22	17.88	2.72	.31
Social Studies	21.54	5.90	45.13	4.74	19.23	3.46	0.00
Biology	2.50	3.13	67.81	.63	25.31	.63	0.00
Chemistry	0.00	2.35	64.12	.29	32.94	.29	0.00
Math	.21	2.29	35.42	1.46	59.06	.94	.63
Physics	0.00	.42	57.92	0.00	41.67	0.00	0.00
Other	10.35	15.44	19.12	5.68	36.91	5.74	6.76
Mean	11.71	4.92	44.45	2.57	33.29	1.97	1.10

In general, respondents (73.11%) felt that more than five hundred questions should be available in a test bank. About half the teachers in Social Studies (52.63%), and exactly half the Biology (50%) teachers, indicated the need for more than 1000 questions in a test bank. Most teachers gave tests containing between 20 and 99 questions (see Table 5), with just over one–third indicating they gave tests with 50 to 99 questions. The teachers tended to give tests approximately once every two weeks (see Table 6).

Table 5

Number of Questions Used in an Average Test

Response	No. of Teachers	Percentage
Less than 10	4	2.06
Between 10-19	19	9.79
Between 20-29	43	22.16
Between 30-39	23	11.86
Between 40-49	32	16.49
Between 50-99	66	34.02
More than 100	7	3.61
Total	194*	100.00

* Note: 198 questionnaires were returned but some contained missing data for certain items.

Table 6

Frequency of Giving Tests

Response	*No. of Teachers*	*Percentage*
once a week	38	19.59
once every two weeks	86	44.33
once every three weeks	46	23.71
once a month	24	12.37
Total	**194***	**100.00**

Question 8 asked, what type of item banking system the teacher was using (or had used). The results (see Table 7), indicated that 63.45% of the teachers have some experience with a test item bank, whether it be manual, on a microcomputer, or on a mainframe computer.

Table 7

Type of Bank Being Used

Response	*No. of Teachers*	*Percentage*
not using one	72	36.55
manual item bank	106	53.80
microcomputer item bank	17	8.63
mainframe item bank	2	1.02
Total	**197***	**100.00**

A marjority of teachers (66.16%) indicated that a computerized test item bank would be very useful (see Table 8). Only two of the 19 teachers presently using a computerized item bank (either on a mainframe computer or microcomputer) did not rate the use of a computerized item bank as *very useful*.

* Note: 198 questionnaires were returned but some contained missing data for certain items.

Table 8

Usefulness of a Computerized Test Item Bank

Response	*No. of Teachers*	*Percentage*
very useful	131	66.16
somewhat useful	43	21.71
limited usefulness	13	6.57
useless	1	.51
don't know	10	5.05
Total	**198**	**100.00**

Over half the teachers (63.45%) indicated that they had used a word processing program on a computer.

2. Ratings of Computerized Test Item Bank Features

Part I: Features Teachers Found Important

To determine which features were either important (score of 3 on the Likert scale) or very important (score of 4 on the Likert scale) for each dimension (Item Structure, Systemization part I, Systemization part II, and Automation), the mean scores and standard deviations for features in each dimension were compared. Means falling just below 3.0 were combined with the corresponding standard error of the mean, to determine if they extended beyond 3.0. The following results were obtained:

1) Figure 6 provides the means and standard deviations for ratings of features in the dimension Item Structure. Table 9 presents those features rated at, or above, a mean of 3.0 in the dimension Item Structure.
2) The dimension Systemization was divided into two parts. Figure 7 presents the mean ratings, +/- one standard deviation, for the features used in examining the item bank. Features rated at or above the mean of 3.0, in part I of the dimension Systemization are presented in Table 10.
3) Part II of the Systemization dimension concerns those features necessary in generating a test. Figure 8 shows a comparsion of the mean ratings, +/- one standard deviation, for those features. Table 11 indicates the mean values for the features in part II of the dimension Systemization which are rated above the mean of 3.0.
4) A comparsion of mean ratings, +/- one standard deviation, for features in the dimension Automation, is presented in Figure 9. The features rated at, or above, the mean of 3.0 are show in Table 12.

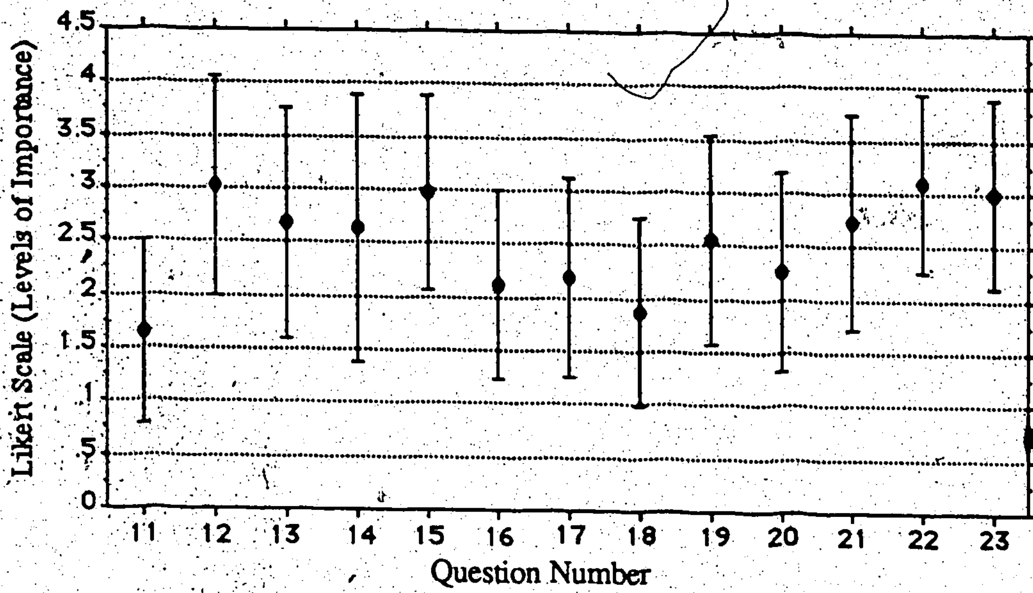

Figure 6. Means +/- 1 Standard Deviations for the Features of Dimension *Item Structure*

Table 9

Questions with Mean Values Above 3.0 in Dimension *Item Structure*

Question #	Statement	Mean
12	Diagram can be included as part of the stem	3.02
15	Availability of character enhancements (e.g., bold, italic, underline, outline, superscript, subscript)	3.02**
22	A series of items linked or associated with one specific item, (e.g., three questions relating to a diagram)	3.10
23	Parallel (similar) items generated by computer	3.05**

** Note: Mean was combined with standard error of the mean

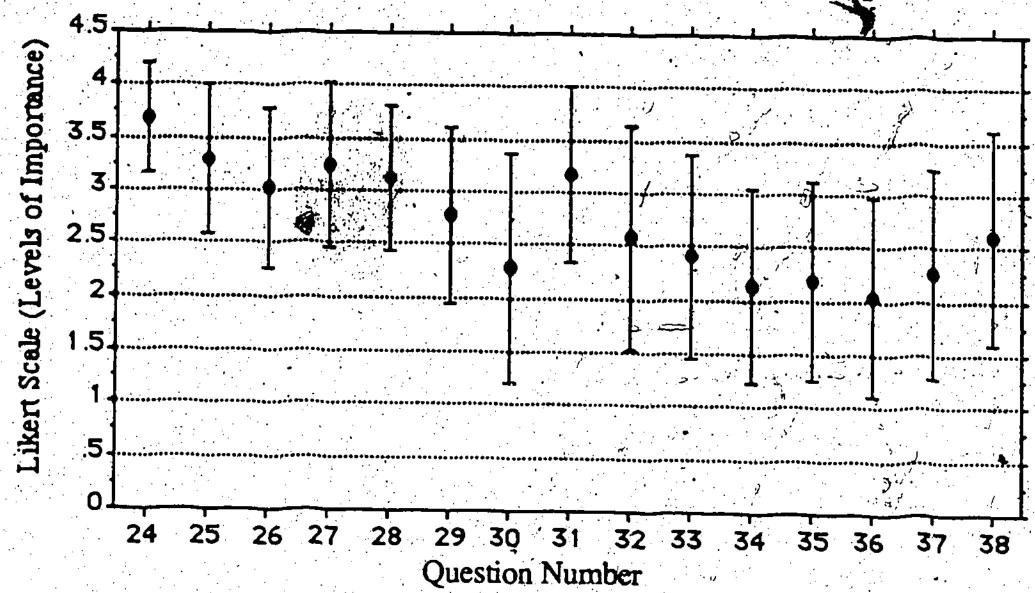

Figure 7. Means +/- 1 Standard Deviations for the Features of Dimension *Systemization Part I*

Table 10

Questions with Mean Values Above 3.0 in Dimension *Systemization Part I*

Question #	Statement	Mean
24	Select by topic (Unit)	3.69
25	Select by instructional objective	3.30
26	Select by cognitive level	3.03
27	Select by item type (e.g., multiple choice, short answer, etc.)	3.24
28	Select by difficulty level	3.12
31	Select by visual inspection of items on computer screen	3.18

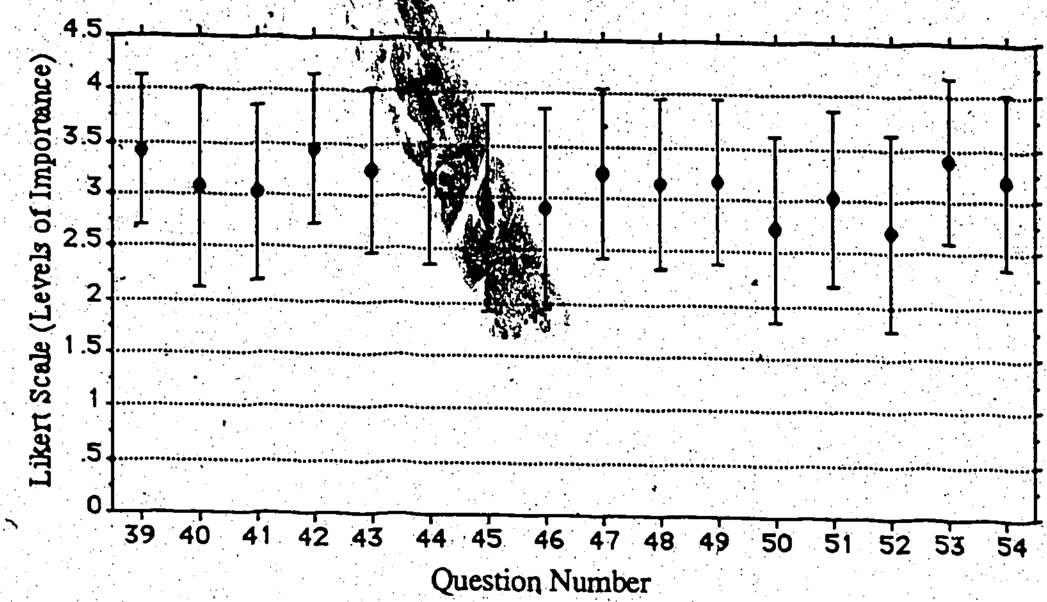

Figure 8 Means and +/- 1 Standard Deviations for the Features in Dimension *Systemization Part II*

Table 11

Questions with Mean Values Above 3.0 in Dimension *Systemization Part II*

Question #	Statement	Mean
39	Easy editing of item text	3.41
40	Easy editing of item text	3.07
41	Select a previously used test	3.04
42	Display a copy of the test before printing	3.44
43	Provision for special directions to be included with different test sections	3.25
44	Generate a set of tests, where different items are used in each of the tests generated but based on same criteria	3.17
47	Keep together items of the same type (e.g., all multiple choice items together)	3.26
48	Keep together items on the same topic	3.14
49	Ability to reorder items before printing	3.18
51	Ability to modify page setup, line spacing and margins	3.04
53	Print answer keys for all versions of the test	3.39
54	Print answer keys for all versions of the test	3.20

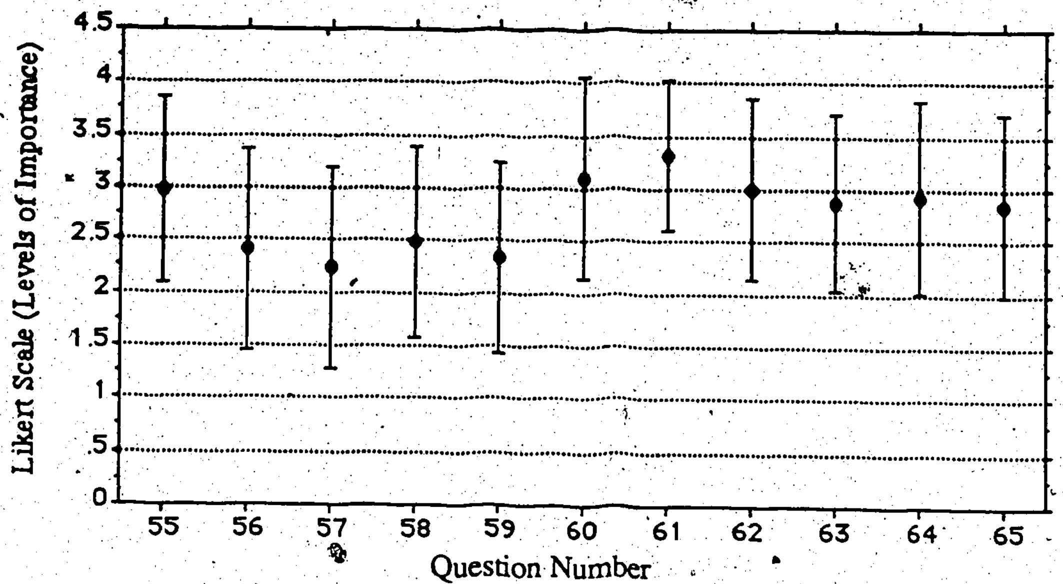

Figure 9. Means and +/- 1 Standard Deviations for the Features in Dimension *Automation*

Table 12

Questions with Mean Values Above 3.0 in Dimension *Automation*

Question #	Statement	Mean
55	A unique identification number should be automatically assigned to each item when it is placed in the item bank	3.05**
60	Security access by password	3.1
61	Difficulty level	3.33
62	Discrimination index	3.06**

** Note: Mean was combined with standard error of the mean

Part II: Proportion of Teachers Responding "Don't Know"

Figures describing the percentage of respondents giving a rating of "Don't Know" to the questions concerning item features (11 to 65) follow. The graphs are set up in exactly the same way with the percentage along the vertical Y-axis and the corresponding question number along the horizontal X-axis. Figure 10 shows the graph of teachers responding "Don't Know" for the questions classified as belonging to the dimension: Item Structure . The percentage of teachers responding "Don't Know" for those questions classified as belonging to part I and part II of the dimension: Systemization are represented in Figures 11 and 12 respectively. Figure 13 represents those teachers responding "Don't Know" for questions classified in the dimension: Automation . Generally, English teachers responded "Don't Know" more often than any other group, as is indicated in all four graphs. In certain questions (29, 30, 35, 36, 55, 62) over 35% of the English teachers selected "Don't Know".

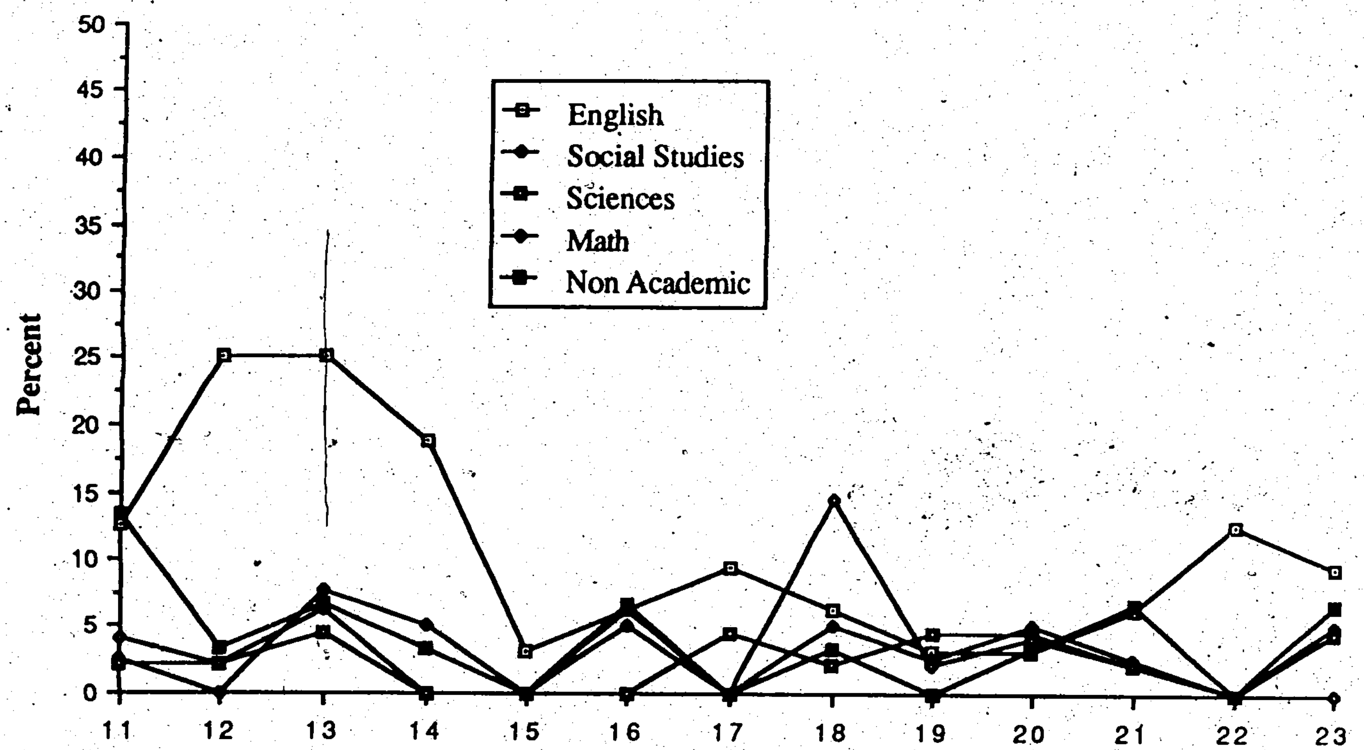

Figure 10. Percentage of Teachers Responding Don't Know to Questions in the Item Structure Dimension.

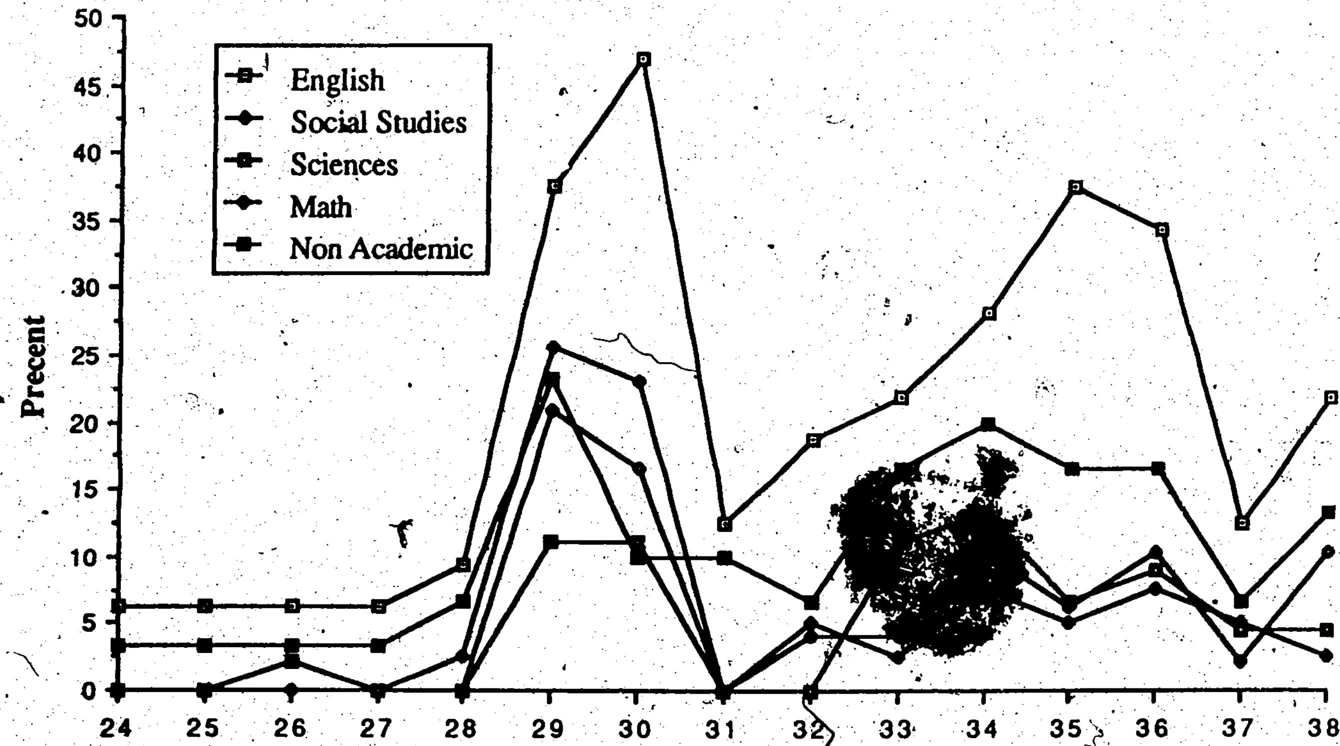

Figure 11. Percentage of Teachers Responding Don't Know to Questions in Part I of the Systemization Dimension

Of the 26 teachers, most simply re-emphasized the importance of some of the features presented in the questionnaire. Specifically, these were as follows: the ability to store test items; easy revision and on-screen editing; randomization of items on a test; randomization of distractors within an item; rearrange questions; generate different tests for the same objectives; security; special symbols; item analysis; graphics; hard copy of bank and ability to have more than one type of question on a test. All these features were mentioned in the survey questionnaire.

Several teachers introduced features that were not present in the questionnaire. One teacher suggested the need for a feature that would convert multiple choice test items into short answer questions or vice versa. Another teacher suggested the inclusion of standardized forms, such as those used in accounting. Another feature suggested for a computerized test item bank is the facility to recognize duplicate questions in the bank. These three features were not mentioned in the survey questionnaire.

The other comments were quite general. They tended to reflect the teachers' enthusiasm for a computerized test item banking system. One teacher suggested that for a computerized item bank to be useful, a large number of items must be available. Two individuals stressed the importance of having a computerized item bank that is compatable with microcomputers presently in the schools. Another teacher identified the fact that many typographical errors are made during data entry. Also, there are often errors in questions related to specific content. One of the most interesting comments was provided by a teacher who had experience with test banking software on a mainframe computer. The teacher stated "I don't like using the mainframe item bank for a couple of reasons, one is the turn around time it takes to get a test and, second, is the lack of control I have in making changes to the items I want to use".

5. Dimensionality of Questionnaire Items

In this section, the questions pertaining directly to features of a computerized test item bank (question 11 to 65) were analyzed using factor analysis techniques. The goal was to explore the validity of the concept of placing items in a three dimensional space. Factor analysis allows these items in the questionnaire to be put into a more manageable form, by reducing the large number of items into "smaller and more comprehensible domains" (Kerlinger, 1979, p. 179). Items which are highly correlated with one another should load more heavily on the same factor. The following three points were considered before the factor analysis was done:

1) Only the 55 questionnaire items (variables) concerned with computerized test item bank features were included.
2) Of the 198 questionnaires returned, only 86 contained complete data for all questions (no missing responses or no "Don't Know" responses). Therefore, only these were included.
3) Cattell's (1966) Scree test for determining the number of factors was used to obtain five factors; these accounted for 48.46% of the total variance.

Table 27 shows the commonalties and five factors for questionnaire items 11 to 65 which resulted from a Varimax rotation. The asterisk indicates the highest factor loadings. Discussion of the results of this factor analysis appears in chapter five part III.

Table 27

Factor Analysis of Questionnaire Item Concerning Features of a Computerized Test Item bank

Question#	Commonalties	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
11	.27	-.07	.36	.16	-.33	-.04
12	.76	.08	.07	.01	-.06	.87
13	.69	.09	.08	.12	.00	.81
14	.45	.04	-.05	.12	-.04	.66
15	.39	.34	-.07	.29	-.04	.37
16	.39	.09	.40	.39	.25	.07
17	.53	-.16	.69	.06	.12	-.03
18	.38	-.02	.61	.10	-.00	-.04
19	.24	.16	.08	.05	.45	-.07
20	.24	-.04	.05	.35	.34	-.08
21	.56	.18	.46	-.08	.51	-.2[illegible]
22	.45	.33	.21	-.24	.32	.3[illegible]
23	.28	.26	.00	-.03	.21	.40
24	.45	.66	-.09	-.07	.09	-.02
25	.41	.15	.02	.10	.56	.26
26	.66	.07	.12	.09	.79	.07
27	.39	.50	.35	.13	.05	.07
28	.59	.14	-.10	.19	.72	.05
29	.59	.26	.14	.26	.66	.06
30	.62	-.02	.51	.26	.54	.01
31	.32	.35	.37	.18	-.05	.17
32	.28	.29	.33	.02	.21	.19
33	.60	.40	.64	.08	-.14	-.09
34	.66	.32	.73	.08	.12	.08
35	.47	.17	.57	.17	.05	.29
36	.60	.15	.72	.16	-.00	.19
37	.48	-.02	.68	.12	.05	-.14
38	.24	.22	.37	.10	.15	.16
39	.55	.59	.04	.12	.26	.34
40	.51	.28	.16	.12	.18	.60
41	.48	.46	.35	.20	.26	.17
42	.48	.66	.04	.09	.12	.13
43	.56	.65	.25	.02	.13	.25
44	.41	.39	.28	.23	.29	.24
45	.47	.46	.22	.45	-.02	-.09
46	.63	.56	.32	.46	.06	-.07
47	.52	.66	.07	.14	.07	.24
48	.50	.68	.07	.02	.19	-.05
49	.62	.71	.11	.23	.24	-.00
50	.39	.49	.19	.22	-.07	.25
51	.44	.49	-.02	.34	-.04	.30
52	.34	.26	.19	.46	.13	.05
53	.45	.64	.05	.05	.14	.13
54	.43	.33	.23	.48	.20	.09
55	.47	-.00	.32	.33	.50	.10
56	.64	.16	.24	.73	.01	.16
57	.67	-.04	.31	.75	.04	.08
58	.60	-.05	.10	.73	.20	.15
59	.51	.01	.31	.66	.02	.11
60	.35	.35	.38	.27	-.05	.09
61	.44	.27	-.07	.52	.18	.25
62	.57	.44	-.11	.49	.32	.11
63	.54	.32	.04	.49	.45	-.05
64	.57	.38	-.03	.56	.31	-.12
65	.56	.44	.11	.50	.27	-.14
	Com. Var.	27.44	20.85	20.85	16.76	14.10
	Total Var.	13.30	10.11	10.10	8.12	6.83

CHAPTER V

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

This chapter is divided in two sections. The first section contains a discussion of the analysis and the second section presents the conclusions and recommendations.

1. Discussion of Results

This discussion closely parallels the presentation of results in chapter four. The first part of this discussion interprets the results that pertain to characteristics of the respondents (Chapter IV, Section one). The second part of this discussion focuses on Sections two, three, and four from Chapter IV, and addresses the purpose of this study. The final part of the discussion evaluates the results obtained from Chapter IV, Section five.

Part I: Characteristics of Respondents

The 56. 86% response rate in this study (see Table 2) is considered fairly good for survey research (Kerlinger, 1973). However, there is a concern that perhaps those teachers who completed the questionnaires felt more positively about computers than those teachers who did not complete the questionnaires. The higher response rate from Mathematics teachers (see Table 3) may have resulted from the fact that these respondents are in a subject area that is closely associated with computers and these teachers have more experience with computers (Vogel, 1985). The total number of questionnaires given to teachers in the individual subject areas is not known. Thus, it is possible that more teachers in Mathematics than in any of the other individual subject areas may have received the questionnaire. The respondents who returned the completed questionnaires represented a fairly experienced group of teachers, with an average of 15.94 years of teaching experience. This indicates a potentially high level of experience in the development of classroom tests. Furthermore, most teachers have had one or more university courses in test construction (61.23%); this indicates an exposure to the methods involved in developing a good test.

Results from the survey indicate that the teachers in the sample have a strong need for a computerized test item banking program (see Table 8). These results are consistent with the study by Brown, Anderson and Louth (1980) which states that most Alberta teachers would welcome computerized help in developing tests. Furthermore, computerized item banks would be of benefit to the teachers surveyed because 63.92% of the teachers administer tests at least once biweekly (see Table 6).

Most teachers use multiple choice and short answer questions; 77.74% of the items used by the teachers in the sample were one of these two item types. When analyzing academic subject areas on an individual basis, over 90.00% of the teachers in Biology, Chemistry, Mathematics and Physics use a combination of multiple choice and short answer types (see Table 4). Over half the Physics teachers (57.92%) and over two thirds of the Biology teachers (67.81%) and Chemistry teachers (64.12%) use multiple choice test items. Unfortunately, quality multiple choice items are often difficult to develop (Chelu & Elton, 1977; Newbould & Massey, 1979). Given that teachers generally use short answer and multiple choice items, and test so frequently, it is unlikely that they develop original tests each time a test is required. As a result, a computerized test item bank, which is well supplied with these types of test items, would make it easier for most teachers to construct relatively unique tests each term.

There seems to be a discrepancy among the teachers sampled concerning the number of items desired in a test item bank. Many teachers indicated they test at least once every two weeks and that they require 20 to 30 items per test (see Table 5), yet they feel that they only need a test item bank containing between 500 and 1 000 items. Constructing tests so frequently would likely require a larger bank because generally several items per objective are required. It is possible that the teachers in the sample did not understand that a test bank must have many more items than the total number of items that would be included in all their tests.

The survey data indicated that the majority of teachers (63.45%) in the sample had some general experience with item banking, but had minimal experience with a computerized item bank (see Table 8). On the other hand, almost two-thirds (63.45%) of the teachers had some experience with word processors. Therefore, teachers in the sample appear to be somewhat familiar with computers. This can be regarded as positive support for the assumption that a majority of teachers are computer literate. Even so, computerized test item bank that is as easy to use as a word processing program would probably have a higher rate of use in the classroom environment than an item bank that did not operate in such a manner.

Part II: Features Teachers Need in a Computerized Test Item Bank

The purpose of this study was to determine which features teachers feel are important for a computerized test item bank. As expected, there were some features that all teachers felt were important and other features which were important only to teachers in certain subject areas. The following is a discussion of the results from the survey, with recommendations for designers of computerized test item banks.

In this survey, item banking features were placed into the three domains described by the Wisconsin group (i.e., Item Structure, Systemization [part I or part II], and Automation). Analysis of the data reveals that there are certain standard features in each of these three domains that the teachers in the sample desire in a computerized test item bank system. Features with mean ratings of 3.0, or greater, were considered to be necessary for inclusion in a test item bank.

Overall, the teachers in the sample found the following features to be important in the Item Structure dimension of a computerized test item bank: diagrams included as part of the stem, character enhancements, a series of items linked to one specific item (e.g., three questions relating to a diagram), and parallel items which are generated by the computer (see Table 16). Closer analysis was used to see if significant differences between subject areas occurred (see Table 17). Question 12, which concerned the need to include diagrams in the item stem, showed that English teachers (mean = 1.625) differed significantly from Social Studies (mean = 2.949), Science (mean = 3.591), Mathematics (mean = 3.383) and Non-Academic teachers (mean = 2.987). Because of the nature of the curriculum material, English teachers expressed little need for a computerized test item bank which stores graphics.

Unlike question 12, which was rated as being important, question 13 (should diagrams be included as part of the alternatives) was not rated as important (see Figure 6). Further analysis revealed a significant difference [$F(4,171) = 15.008$, $p < .0001$] between the subject areas; English teachers (mean = 1.542) were also responsible for lowering the overall rating of question 13 to less than 3 (see Table 18). For example, Mathematics (mean = 3.222) and Science (mean = 3.023) teachers found this to be a important feature. Mathematics and Science teachers often use graphs in the stem or in the alternatives. Biology teachers also use many diagrams in their exams. As one Biology teacher commented, "I would like to stress the importance of being able to include and/or insert graphics into test items stored".

Currently, very few systems allow for the inclusion of graphics. According to Hiscox (1984), this is a major drawback of most microcomputer test item bank programs and the problem usually stems from inadequacies in the software and hardware. To print graphics on paper, especially the "camera-ready" quality needed for high-quality test production, is costly. One of the few microcomputer test banking programs reviewed which could handle graphics was measureUP. This program uses built-in graphic routines, a high resolution screen, and, if a Laser Writer Printer is available provides near "camera ready" printing. This compares favorably with larger, more expensive systems, such as the one developed for the State of Wisconsin (Burke, Kaufman & Webb, 1985). Software designers contemplating the development of new computerized test item banks are well advised to include a graphics facility in their item banking program.

Teachers in all of the academic subject areas rated character enhancements as being important (see Table 10). Teachers often find it frustrating to incorporate different character fonts into their exams. Deck (personal communication, July, 1987) indicated that this was one of the major problems experienced by the Northwest Regional Educational Laboratory in developing the Utah State Exams. Very few microcomputer test item banking systems provide character enhancement features (Nitko & Hsu, 1984). Therefore, character enhancement should be included in the development of any future computerized test item bank.

Even though character enhancements received a high rating, Figure 6 indicates special character sets (question 14), did not. A more detailed analysis however, revealed that significant differences [$F(4,180) = 47.179$, $p < .0001$] existed between teachers in different subject areas (see Table 19). The low mean ratings by English teachers (1.462) and Social Studies teachers (1.703) lowered the overall mean rating. This feature is likely important for both Mathematics teachers (mean = 3.75) and Science teachers (mean = 3.2) because they rely on special character sets, such as Greek symbols. The availability of special character sets is a feature that is missing in most word processors used by teachers. Open ended responses indicated that Mathematics and Science teachers have often been frustrated by the lack of such a feature (Appendix C).

All groups rated the ability to link items as being important (see Table 10). This was expected, because teachers in all subject areas usually have several questions that are related to a passage, diagram, or table. Even though this is an important feature, few of the microcomputer test bank systems which were reviewed had such a capability. This feature should also be incorporated in any new system being developed.

Computer generation of parallel items is regarded as an important feature by teachers from all subject areas (see Table 10). This feature could help to reduce the amount of time teachers require to build items for an item bank (Millman & Outlaw, 1978). Presently, computer generated parallel items are based on an algorithm and meant for questions in mathematics (Millman, 1984), and chemistry (Johnson, 1981). Denney (1973) describes the Question Pool Management System which generates multiple choice test items from a series of stems, correct alternatives and wrong alternatives. Denney states that, if the stem has seven correct alternatives and seven wrong alternatives associated with it, 245 multiple choice questions containing five alternatives with only one correct answer could be generated. Unfortunately, Hsu and Sadock (1985) state that no microcomputer programs exist which generate questions either by algorithm or by Denney's method. This is probably due to the limited storage space these systems currently have available (Hsu & Sadock, 1985). With higher capacity hard disk drives being developed for microcomputers, there is a greater potential for easier development of more sophisticated computerized test item bank programs which allow for the generation of parallel items.

Two features that were not rated as being important, but showed a significant difference between the mean ratings of teachers in the subject areas were, question 19 (assigning weights to items) (see Table 19), and question 21 (grouping items by test type) (see Table 20). In both cases, English teachers rated these two statements as important; this differed significantly from the ratings given by Social Studies, Mathematics, and

Science teachers. Perhaps English teachers view the ability to assign weights as being more important because they tend to assign more essay type questions (47.34%) than do teachers in the other subject areas. It is important that students know the value of a question so that they can devote more time to items that are worth more if they so desire.

Teachers in the sample found the following features of the dimension Systemization Part I to be important: select by topic, instructional objective, cognitive [illegible] item type, difficulty level, and visual inspection of items on the screen (see Table [illegible] Test construction courses are likely to stress the importance of selecting items according to difficulty level; the majority of teachers surveyed have completed some course work in this area and this could account for this feature achieving a high rating. It was assumed that teachers would rate the ability to view the items selected by certain criteria, before placement in the examination, as being important, because the item may not be appropriate for a particular academic group of students. It appears that all the above features should be standard features in a test banking system.

Although the teachers did not rate question 26 (select items by cognitive level) and question 28 (select by item difficulty level) as being important (see Figure 7), these questions showed a significant difference between subject areas (see Table 22 and Table 23 respectively). In both cases non-academic subject area teachers differed significantly from teachers in English. Possibly, teachers in the non-academic subject areas are more concerned with evaluation of motor skills, such as typing 30 words per minute.

Teachers in all subject areas found many of the features in the Systemization Part II dimension to be necessary (see Table 12 and Figure 8). These features dealt primarily with text formatting. Features such as easy editing of item text, easy editing of item graphics, displaying a copy of the test before printing, ability to modify page setup, and line spacing and margins are characteristics of word processor and graphics programs that need to be incorporated into a test item bank system. Features, such as the provision of special directions to be included in different sections of the test, keeping together items of the same type, and reordering of items before printing, are useful in preparing and organizing a test. Generating different test versions goes hand-in-hand with generating answer keys for all the test versions. This capability of generating multiple versions of a test is important for those teachers who teach more than one section of the same course. Often a test is written by one class one day and another class takes a similar test the following day. Since, some students who have not yet written the test will possibly discuss the contents of the test with those who have already written the test, this could give the students who write the test at a later time an unfair advantage over those who write the test earlier.

The ability to merge different test banks is another potentially important feature because this permits schools to more easily and economically share test items. "Test quality may be readily improved by encouraging and promoting collaboration in assessment" (Stiggins and Bridgeford, 1985, p. 284). Furthermore, existing items from other areas would be available for inclusion in the tests. For example, the University of Florida offers access to 10 000 validated science questions via phone modem (Dawson, 1987) and Tescor, Inc. is storing validated test items on CD ROM for distribution (Miller, 1987). Merging item banks can reduce the costs involved with item development and provide teachers with a larger, more varied pool of items.

In the dimension Systemization Part II, only one feature, easy editing of graphics, shows a significant difference in the importance assigned by the different subject areas (see Table 24). English teachers gave the graphics capability a low rating of importance (mean = 2.308), whereas all other subject areas found it important. This is similar to the response given by English teachers to earlier questionnaire items (questions 12 and 13) regarding the ability to include diagrams as part of the stem and alternatives.

Only four of the 11 features in the Automation dimension were considered as important by the teachers in the sample (see Table 13). Assigning a unique identification number to each test item was considered important, probably because this allows for easy management of the item bank. Items can be retrieved for viewing by identification number and then, if necessary, easily modified. Security access, which is limited by a password, was also given a high rating. This comes as no surprise, considering the potential damage that could be done if unauthorized users gained access to the item bank. Because many items can be stored in a test bank, the security of items must be maintained. This is a particularly important issue (Muiznieks & Dennis,1979), although many writers feel that a large test item bank would allow secrecy to be done away with (Davis & Zacharias, 1982; Millman & Arter, 1984).

The teachers were asked to indicate which type of statistical information should be maintained if the questions were scored by a machine. A system that automatically calculates item difficulty and an item discrimination was considered important by the teachers (see Table 13). Yet, when teachers were asked whether they wanted to select items based on the discrimination index (question 29), it was rated below 3.0 (see figure 7). Although most of the teachers had at least one test construction course, they did not give high ratings (see Figure 9) to features such as, distractor information (question 63), provision of means for high, median, and low groups (question 64), and number of times a response to each item was omitted (question 65). In fact, little difference was found between teachers with, and those without, some formal instruction in test construction (see Table 25). According to Gullickson (1984), "most teachers have had at least one course in education measurement, but most also have a limited knowledge in measurement" (p. 248). If computerized test item banks are to be used to improve evaluation methods, teachers will need to understand and make use of such features for test and item analysis.

Teachers in the English subject areas generally responded "Don't Know" more often then teachers from any of the other subject areas (see Figures 10 to 13). This may indicate a lack of certainty on the part of English teachers regarding the importance of features in computerized test item bank. The high percentage of "Don't Know" responses from English teachers may also be related to the fact that many of the questions in the survey refer to multiple choice test items.

It was expected that if teachers wanted the software of a computerized test item bank to calculate item discrimination indices, that these teachers would also select items based on the discrimination indices. The discrepancy between means ratings for question 29 (select by discrimination index) and question 62 (system automatically calculates discrimination index) might be accounted for by the high percentage of "Don't Knows".

Of the 19 teachers with experience in using a computerized item bank, 17 indicated a computerized test item bank would be very useful. Further analysis revealed significant differences [$F(2,185) = 7.197$, $p < .001$] between: 1) non-users of a test item bank, 2) users of a manual test item bank, and 3) users of a computerized test item bank. Table 24 indicates that teachers *who have not used an item bank* (mean = 3.364) would find a computerized test item bank less useful than teachers who have used manual test item banks (mean = 3.689) or teachers who have used computerized test item bank (mean = 3.895). It is assumed that those using a manual bank will find that a computerized test item bank will reduce the tedious tasks involved in developing a test.

Several teachers who indicated they were using a microcomputer test banking system, were only using a word processing program (i.e., Appleworks) (see Table 26). The results also indicate that four teachers were not sure of the type of computerized item bank program they had used and, therefore they specified the computer system. The fact that a number of teachers have used word processing programs indicates the features generally present in word processing programs can assist a teacher who is developing a test. Unfortunately, such word processing features represent a primitive version of the capabilities available in a sophisticated computerized test item banks.

Part III: Validity of Questionnaire Items

As noted in the literature review, a computerized test item bank can be defined in terms of the three dimensions outlined by the Wisconsin model (i.e., Item Structure, Systemization, and Automation). These item bank dimensions became the theoretical basis upon which the items for the questionnaire used in this study were designed. A confirmatory factor analysis was carried out on the item response data to determine if the questionnaire items could be factored into the three dimensions underlying the Wisconsin model, thereby establishing factorial validity.

When the item data were factor analyzed and the Scree test (Cattell, 1966) applied (see Table 27) five main factors were revealed. Factor one accounted for 13.30% of the total variance and substantiates the dimension of Systemization (Part II) which includes features generally used in creating a test; two other items (number 24 and 27) also loaded heavily on factor one. Both of the items were deemed to be important features in a test item bank (see Figure 7). It is interesting to note that question 24 had the highest mean rating of any of the features in factor one.

The factor analysis also revealed that the dimension Systemization (Part I) was represented by two factors (factor two and factor four). Factor two accounted for 10.11% of the total variance; it was composed of items 11, 16-18, 31-38, and 60. These items represent general item bank features which teachers would use for selecting and accessing items for a test. Factor four consisted of items 19, 21, 25, 26, 28, 29, 30, and 55 and accounted for 8.12% of the total variance. These eight questionnaire items are concerned with specific item bank features which teachers would require when selecting items for inclusion in a test.

Factor three accounts for 10.10% of the total variance and contains items 20, 52, 54, 56-59, 61-65; this factor supports the dimension of Automation. Factor five accounted for 6.83% of the total variance and contains items 12-15, 22, 23, and 40. This factor contains item related to graphics.

Overall, the factor analysis showed support for the dimensions within which the questionnaire items were placed when the questionnaire was designed. This indicates that the Wisconsin model acted as a good theoretical basis upon which to develop the questionnaire. As a result the questionnaire shows a high level of factorial validity.

2. Conclusion and Recommendations

The purpose of this study was to determine the features that high school teachers regard as being are important for inclusion in a computerized test item banking system. The questionnaire developed for this study identified 55 features of a computerized test item banking system which may be beneficial to the classroom teacher. The ideal computerized test item bank would incorporate all these features; unfortunately such a system does not exist. As a result, the questionnaire used in this study provided a means by which classroom teachers rated item banking features according to levels of importance.

The results of this study should provide schools, subject area departments, and individual teachers, who are looking for a test item banking system, with a reference point from which they can identify a test item banking system which is most applicable to their situation. For example, a school may want to purchase a test item banking system which contains features that every teacher in the school can use. On the other hand, the science department in a school may want an item bank that is designed specially to handle graphics or has special character enhancements.

The implementation of a computerized test item bank for the classroom environment will not occur simply because adequate software is provided. In this study, a high percentage of teachers responded "Don't Know" to the items and over 43% of the teachers failed to respond to questionnaires that were issued. This may indicate a certain lack of awareness regarding computerized test item banking. Inservice programs might help increase teachers' awareness of the process of using computerized test item banking and of the benefits of using such a system. To help achieve this awareness requires a commitment from school administrators, local school boards, university departments of education and the provincial government.

Pre-service teacher training at the university level should include the practical application of computer technology and software to the development of classroom tests. Gullickson and Hopkins (1987) surveyed 28 institutions that grant baccalaureate degrees in the midwestern U.S. and found "the use of a computer for test scoring, item analysis, or test development was rarely mentioned" (p. 14) to education students. Student teachers who are taught to use a computerized item bank will possibly feel more comfortable using such a tool in the test development process. Furthermore, practicing teachers also require further training in educational measurement techniques (Gullickson & Hopkins, 1987).

It is difficult for individual teachers to find the time which is necessary to construct a comprehensive test item bank in any subject area. Therefore it is essential for teachers to cooperate and to combine their efforts to develop a bank of items. Teachers should also be able to include quality test items which are already available from other sources. Presently Alberta Education maintains a large number of test items which were designed for previous provincial diploma examinations. These test items could easily form a large test item bank. With this large base of items, teachers can begin to add their own items and build their own banks which would have a sense of uniqueness.

The results of this study, and some of the limitations already identified, lead to several suggestions for future research. First, a larger more representative sample of teachers should be obtained and this study replicated. Second, this study did not address the needs of teachers in elementary school, junior high school or postsecondary institutions. All these groups should be surveyed and the results compared to determine if different needs exist between groups. Third, a study to investigate the procedures followed by teachers in developing tests and how these procedures can be improved through computerized test item banking systems should be undertaken. This would involve an assessment of how much teachers already know about test development, item construction, and item analysis. Finally, a study comparing teachers in schools that use and in schools that do not use computerized test item banks could provide valuable information on how the testing processes differ.

In summary, this study identified a number of features of computerized test item banks that are most important to high school teachers, in general, and those features which are important to high school teachers in different subject areas. An underlying structure to these features was formulated and five major factors identified. It was revealed that high school teachers feel that computerized test item banking is a useful tool.

REFERENCES

Alberta Curriculum Guides, Published by Alberta Education.

Alberta Education. (1985). Computer Courseware Evaluations: January, 1983 to May, 1985, Edmonton, Alberta: Author

Anderson, J. O. (1987, June). Teacher Practices in and Attitudes Towards Student Assessment. Paper presented to the annual meeting of the Canadian Educational Researchers' Association, McMaster University, Hamilton, Ontario.

Arter, J. A. & Estes, G. D. (1985). Item Banking for Local Test Development: Practitioner's Handbook. Portland: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 266 166)

Augarten, S. (1985). Bit by Bit. New York: Ticknor & Fields.

Baker, F. B. (1971). Automation of test scoring, reporting and analysis. In R. L. Thorndike (Ed.) Educational Measurement, (2n ed.) Washington, DC: American Council on Education.

Baker, F. B. (1984). Technology and Testing: state of the art and trends for the future. Journal of Educational Measurement, 21(4), 399-406.

Baker, F. B. (1986). Item Banking in Computer-Based Instructional Systems. Applied Psychological Measurement, 10(4), 405-415.

Borg, W. R. & Gall, M. D. (1983). Educational Research: an introduction (4th ed.). White Plains, New York: Longman.

Bowers, J.J. (1986, April). A Microcomputer-Based System to Manage Health Professions Test Development and Item Banking Activities. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California.

Bradbard, D. A. (1986). A System for Generating Classroom Tests. Journal of Educational Technology Systems, 14(3), 261-268.

Broduer, D. R. (1986). Test Generator Program Features That Facilitate Classroom Testing. Educational Technology, 26(11), 39-42.

Brown, B. (1982). Automated test and quiz production. Classroom Computer News, 2(4), 33-35.

Brown, A., Anderson, L., & Louth, D. A. (1980). A practical approach to computer based instructional support. A report submitted to Alberta Education, Edmonton, Alberta: Alberta Education.

Brzezinski, E. J. (1984). Microcomputers and Testing: where are we and how did we get there? Educational Measurement: Issues and Practices, 3, 7-10.

Brzezinski, E. J., Powers, L. L., Tamura, S. G., & Hiscox, M. (1980). Part 3: Results of Item Bank Telephone Survey, Portland: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 205 027)

Buck, G. (In press). M. E. LaZerte (1885-1975): Educational Innovator. Canadian Journal of Education.

Burke, N. W., Kaufman, D. B., & Webb, N. (1985). The Wisconsin Item Bank. In G. D. Estes (Ed.), Examples of Item Banks to Support Local Test Development: Two Cases Studies with Reactions, Porland: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 266 168)

Byrne, C. J. (1976). Computerized Question Banking System: I - The State of the Art. British journal of Educational Technology, 2(7), 44-64.

Carter , K. (1984). Do teachers understand the principles for writing tests? Journal of teacher Education, 35(6), 57-60.

Cattell, R. B. (1966). The Scree Test for the Number of Factors. Multivariate Behavioral Research, 1(2), 245-276.

Chelu, C.J. & Elton, L.B. (1977). An item bank for multiple choice questions, Physics Education, 12, 263-266.

Collins, M. M., Ellis, A. P., Fiske, J. M., & Genco, H. L. (1985). Computer Assisted Test Bank. Journal of Nursing Education, 24(8), 349-350.

Curtis, H.A. (1972). The Development and Management of Banks of Performance Based Test Items, Symposium presented at the Annual Meeting of the National Council on Measurement in Education, Chicago. (ERIC Document Reproduction Service No. ED 072 099)

Davis, S. & Zacharias, J. (1982). Testing In the Service of Education. Classroom Computer News, 2(4), 28-32.

Dawson, G. (1987). 10, 000 Science Questions by Modem. The Science Teacher, 54(3)

Deck, D., & Estes, G. D., (1984, April). Microcomputer software support for item banking. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Deck, D., Nickel, P., & Estes, G. (1985). Reviews of Microcomputer Item Banking Software, Portland: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 266 167)

Denney, C. (1973). There is more to a test pool than data collection. Educational Technology, 13(3), 19-20.

Dudley-Marling, C. & Owston, R. D. (1987). The State of Educational Software: A Criterion-Based Evaluation. Educational Technology, 27(3), 25-33.

Lippey, G. (Eds.). (1973). Computer Assisted Test Construction [Special Issue]. Educational Technology, 13(3).

Ebel, R. L. (1972). Essentials of Educational Research Measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Enger, R. C. (1981). The USAF Academy basic physics exam file. Journal of College Science Teaching, 10(4), 240-243.

Estes, G. D. (Ed.). (1985). Examples of Item Banks to Support Local Test Development: Two Cases Studies with Reactions, Portland: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 266 168)

Estes, G. D., & Arter, J. A. (1984). Item Banking for State and Local Test Development and Use: Specific Guidelines for Local practitioners, Portland: Northwest Regional Educational Laboratory, (ERIC Document Reproduction Service No. ED 242 763)

Fremer, J. (1984). Using Microcomputers to Develop Test: an alternate point view. Educational Measurement: Issues and Practices, 3, 14-15.

Gronlund, N. E. (1985). Measurement and Evaluation in Teaching. (5th ed), New York: Macmillan.

Green, K. E. & Stager, S. F. (1986, April). Effects of Training, Grade Level, and Subject Taught on the Types of Tests Items Used by Teachers. Paper presented at the National Council on Measurement in Education Annual Meeting, San Francisco, California.

Gullickson, A. R. (1982). The Practice of Testing in Elementary and Secondary Schools. (ERIC Document Reproduction Service No. ED 229 391)

Gullickson, A. R. (1984). Teacher Perspectives of Their Instructional Use of Tests. Journal of Educational Research, 77(4), 244-248.

Gullickson, A. R. (1986). Teacher Education and Teacher-Perceived Needs in Educational Measurement and Evaluation. Journal of Educational Measurement, 23(4), 347-354.

Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice, 4(1), 15-18.

Gullickson, A. R., & Hopkins, K. D. (1987). The Context of Educational Measurement Instruction for Preservice Teachers: Professor Perspectives. Educational Measurement: Issues and Practice, 6(3), 12-16.

Haertel, E. (1986, April). Choosing and Using Classroom Tests: teachers' perspectives on assessment. Paper presented at the meeting of ~~the~~ American Educational Research Association, San Francisco, California.

Hambleton, R. K. (1984). Using Microcomputer to Develop Tests. Educational Measurement: Issues and Practices, 3, 10-14.

Hambleton, R. K., Anderson, G. E., & Murray, L. N. (1983). Applications of Microcomputers to Classroom Testing, In W. E. Hathaway, (Ed.), Testing in the Schools: New Directions for Testing and Measurement, San Francisco: Jossey-Bass.

Hathaway, W., Houser, R., & Kingbury, G. (1985). A Regional and Local Item Response Theory based on Test Item bank System. In G. D. Estes (Ed.), Examples of Item Banks to Support Local Test Development: Two Cases Studies with Reactions, Portland: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 266 168)

Hazlett, C. B. (1973). MEDSIRCH; Multiple choice test items. Educational Technology, 13(3), 24-25.

Hiscox, M. D. (1983). A Balance Sheet for Educational Item Banking. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Quebec. (ERIC Document Reproduction Service No. ED 235 209)

Hiscox, M. D. (1984). A planning guide for microcomputers in educational measurement. Educational Measurement: Issues and Practice, 3, 28-34.

Hsu, T. C. (1986, April). Developments in microcomputer applications to testing. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California.

Hus, T. C., & Nitko, A. J. (1983). Microcomputer testing software teachers can use. Educational Measurement: Issues and Practice, 2, 15-18, 23-30.

Hsu, T. C., & Sadock, S. F. (1985). Computer-Assisted Test Construction: The State of the Art Office of Educational Research and Improvements (ED), Washington, DC. (ERIC Document Reproduction Service No. ED 272 515)

Johnson, J. K., Willis, Van., Seely, O. Jr., & Moore, J. W. (1981). Computer-Assisted Test Construction in Chemistry. Journal of Chemical Education, 58, 177-181.

Johnson, S. & Maher, B. (1982). Monitoring Science Performance Using a Computerized Question Banking System. British Journal of Educational Technology, 12(13), 97-106.

Ju, T. P. (1985). Sample of user's Evaluation Form. In T. C. Hsu, & S. F. Sadock, Computer-Assisted Test Construction: The State of the Art Office of Educational Research and Improvements (ED), Washington, DC. (ERIC Document Reproduction Service No. ED 272 515)

Kanter, H. M: (1985). Computer Applications of Educational Measurement Concepts, New York: Macmillan.

Kerlinger, F. N. (1973). Foundations of Behavioal Research. (2nd ed). New York: Holt, Rinehart & Winston.

Kerlinger, F. N. (1979). Behavioral Research: a conceptual approach. New York: Holt, Rinehart & Winston.

Kolmes, J. A. (1983). The second debut for provinical exams. The Canadian School Executive, 2(9), 5-9.

Kotz, S., & Johnson, N. L. (1982). Encyclopedia of Statistical Sciences. (Vol. I). New York: John Wiley & Sons.

Legg, S. M. (1982). The use of precalibrated item bank to establish and maintain cutoff scores. A case study of the Florida Teacher Certification Examination. Paper presented at the annual meeting of the National Council on Measurement in Education, New York. (ERIC Document Reproduction Service No. ED 221 570)

Lippey, G. (1973). The Computer can Support Test construction in a Variety of Ways. Educational Technology, 13(3), 10-12.

Lippey, G. (Ed.). (1974). Computer Assisted Test Construction. Englewood Cliffs, NJ: Education Technology Publications.

Mason, G., Mcpherson, B., Hum, D., Roberts, L., & Anderson, A. (1983). Survey Research Methods. Institute for Social and Economic Research, University of Manitoba: Faculty of Arts.

McLean, L. D. (1985). The Craft of Student Evaluation in Canada. A Report to the Canadian Education Association on policies, practices and uses of assessments of student achievement. Canadian Education Association, Toronto., Ontario.

Miller, D. C. (1987). CD ROM Joins the New Media Homesteaders. Educational Technology, 27(3), 33-35.

Millman, J. (1984). Individualizing test construction and administration by computer. In R. A. Berk (Ed.), A Guide to Criterion-Reference Test Construction. Baltimore, Maryland: John Hopkins University Press.

Millman, J., & Arter, J. A. (1984). Issues in Item Banking. Journal of Educational Measurement, 4(21), 315-330.

Millman, J., & Outlaw, S. W. (1978). Testing by Computer. Journal of the Association for Educational Data Systems, 11(3), 57-72.

Moore, J. W., Prosser, F., & Donovan, G. (1974). Computer Generated Repeatable Tests in Chemistry. In J. S. Mattson (Ed.), Computers in Chemistry and Instrumentation, New York: Marcel Dekker.

Muiznieks, V., & Dennis J. R. (1979). A look at Computer-Assisted Testing Operations: The Illinois Series on Educational Application of Computers. New York: Exxon Education Foundation. (ERIC Document Reproduction Service No. ED 183 191)

National Commission on Excellence in Education. (1983). A Nation at Risk: The Imperative for Educational Reform. The Elementary School Journal, 84, 113-129.

National Governors' Association. (1986). Time for Results: The Governors' Report on Education. Washington, DC: (ERIC Document Reproduction Service No. ED 279 603)

Nelson, L. R. (1984). Using Microcomputers to Assess Achievement and Instruction. Educational Measurement: Issues and Practice, 3, 22-26.

Newbould, C. A., & Massey, A. J. (1977). A Computerized Item Banking System (CIBS). British journal of Educational technology, 8(2), 114-123.

Nitko, A. J. (1983). Educational tests and measurement: An introduction. New York: Harcourt Brace Jovanovich.

Nitko, A. J., & Hsu, T. C. (1984). A Comprehensive Microcomputer System for Classroom Testing. Journal of Educational Measurement, 21(4), 377-390.

Peterson, J. A., & Meister, L. L. (1983). Managing a Test Item bank on a Microcomputer: Can it help you and your Student. T.H.E. Journal, 20, 120-122.

Petruk, M. (1986). Microcomputers in Alberta Schools. Media and Technology Branch: Alberta Education

Pine, J. A. (1986, January). Faculty Ratings of Test Bank Program Features. Paper presented at the Association for Educational Communications and Technology, Las Vegas, Nevada.

Pressey, S. L. (1926). A simple apparatus which give tests and scores—and teaches. School and Society, 23(585), 373–376.

Pressey, S. L. (1927). A machine for automatic teaching of drill material. School and Society, 25(645), 549–552.

Romaniuk, E. W. (1980). Computerized Item Banking Systems. A report submitted to Alberta Education: Province of Alberta.

Seely, O., & Willis, V. (1976). SOCRATES' test retrieval at the California State University and Colleges. AEDS Journal, 9(3), 65-70.

Skinner, B. F. (1958). Teaching Machines. Science, 128(3330), 969-977.

Stiggins, R. J., & Bridgeford (1982). Final research report on the nature, role and quality of classroom performance assessment. Portland: Northwest Regional Educational Laboratory.

Stiggins, R. J. & Bridgeford (1985). The Ecology of Classroom Assessment. Journal of Educational Measurement, 22(4), 271-286.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom Assessment: A Key to Effective Education. Educational Measurement: Issues and Practice, 5(1), 5-17.

Stodola, Q. (1965). Data Processing Procedures to Improve Classroom Testing. Educational and Psychological Measurement, 25(3), 885-886.

Stodola, Q. (1973). Use of Computer Assembled Tests in the California State University and College System. Educational Technology, 13(3), 40-41.

Stodola, Q. (1974). Item Classification and Selection. In G. Lippey (Ed.), Computer Assisted Test Construction. Englewood Cliffs, NJ: Education Technology Publications.

Thiel, C. C., & Zwarts, M. A. (1986). Development of a testing Service System. Applied Psychological Measurement, 10(4), 391-403.

Toggenburger, F. (1973). Classroom Teacher Support System. Educational Technology, 13(3), 42-43.

Vogel., G. (1985). A Microcomputer-Based System for Filing Test Question and Assembling Examinations. Journal of Chemical Education, 62, 1024-6.

Ward, W. C. (1984). Using microcomputers to administer tests. Educational Measurement: Issues and Practice, 3, 16-20.

Weber, M. (1984, April). Item Banking: New Technologies, Old Issues. Paper presented at the Annual Meeting of the American Educational Research Association.

Wright, B. D., & Bell, S. R. (1984). Item Banks: What, Why, How. Journal of Educational Measurement, 4(21), 331-345.

Zacharias, J. R. (1979). The People Should Get What They Need: An optimistic view of What Tests Can Be. National Elementary Principal, 58(2), 48-51.

Appendix A

Survey Letter to Teachers

Dear Teacher:

The attached survey is intended to assess the needs of classroom teachers regarding computerized test item banks. The results of this survey will help provide some of the information which is necessary for selecting an appropriate test item banking system. Presently many different types of computerized test item bank programs are available and a problem exists as to which program best suits the needs of classroom teachers. Therefore I am particularly interested in obtaining your responses to this survey. Your experience in classroom testing should contribute significantly toward solving some of the problems in this area of education. Although you may not be familiar with computers, you are likely familiar with the concept of a test item bank. In its simplest form, the term test item bank refers to *a large group (bank) of easily accessible test items.* The goal of any test item bank is to minimize teacher time in constructing quality tests. Test banks do not have to be computerized, however the use of computers allows the process of item banking to be completely automated and thereby reduce the time needed to construct tests.

The survey is divided up into the following three sections:

1) Item Structure - which is concerned with the attributes to be included with each item.
2) Systemization - refers to the process whereby items are selected and tests are generated.
3) Automation - automatically provides many features which normally are done manually.

The enclosed survey has already been pilot tested with a group of teachers, and revised. An average of 15 minutes is required to complete this survey. Please complete the enclosed survey by June 8, 1987 and return it to your school's General Office. Any comments that you may have concerning any aspect of the survey or of computerized test item banks are welcome. Space has been allocated for this at the end of the survey.

In appreciation of your help I am offering teachers involved in this study a demonstration of one of the latest microcomputer test item banking systems at 4:00 pm on **June 4, 1987, in room 2-115, Education North.** There is no cost to attend. For more information call 432-3762.

Thank you for your time,

Michael Carbonaro

Appendix B

Questionnaire

General Information:

Please place a checkmark beside the information that applies to you, unless otherwise specified.

1) How many years have you been teaching?

____

2) How many university courses in test construction have you taken?
____ none
____ 1
____ 2
____ 3 or more

3) What subject do you **primarily** teach? (***Please check only one***)
____ English
____ Social Studies
____ Biology
____ Chemistry
____ Math
____ Physics
Other (Specify)____________

4) Which types of test items do you use? *(Please indicate %, e.g. 80% Essay, 20% Multiple choice, total 100%)*
____ % Essay
____ % Fill in the blank
____ % Multiple choice
____ % Matching
____ % Short answer / workout problems
____ % True/False
Other (specify %)____________

5) How many questions would you need in an item bank for your subject area?
____ 250 or less
____ 251- 500
____ 501-1000
____ 1000 or more

6) How many questions do you use on an average test?
____ Less than 10
____ 10-19
____ 20-29
____ 30-39
____ 40- 49
____ 50-99
____ 100 or more

7) How often do you give tests?
____ once a week
____ once every two weeks
____ once every three weeks
____ once a month

8) What type of system are you using (or have used)?
____ not using one
____ manual item bank
____ microcomputer item bank
____ Mainframe item bank

If using a microcomputer item banking program please give the name below,

____

9) How useful would a computerized test item bank program be to you?
____ Very useful
____ Somewhat useful
____ Limited usefulness
____ Useless
____ Don't know

10) Have you ever used a word processing program?
____ Yes
____ No

NOTE: Many of the following questions in this survey refer to multiple choice test items.

Item Structure:

In a computerized test item bank, questions are usually selected by teachers according to specific attributes such as course, unit, subject, objective, cognitive level, etc. Below is a list of attributes that can be entered either as part of the item itself or in association with the item. Please rate the importance to your subject area of the following attributes.

	Very Important	Important	Somewhat Important	Not Important	Don't Know
11) Author or source of item	___	___	___	___	___
12) Diagram can be included as part of the stem	___	___	___	___	___
13) Diagrams can be included as a part of the alternatives	___	___	___	___	___
14) Availability of special character sets (symbols, e.g. $\sum$)	___	___	___	___	___
15) Availability of character enhancements (e.g. **bold**, *italic*, underline, outline, superscript, subscript)	___	___	___	___	___
16) Individual teacher comments about an item	___	___	___	___	___
17) Cross reference of items to specific pages in text books	___	___	___	___	___
18) Cross reference of items to audio [illegible] materials	___	___	___	___	___
19) Assignment of teacher specified weights to items (how many marks is the item worth)	___	___	___	___	___
20) Average completion time per item	___	___	___	___	___
21) Items grouped according to the intended type of test (e.g. quiz, unit exam, final)	___	___	___	___	___
22) A series of items linked or associated with one specific item, (e.g. three questions relating to a diagram)	___	___	___	___	___
23) Parallel (similar) items generated by computer	___	___	___	___	___

Systemization:

Test banks usually allow items to be retrieved according to attributes attached to each item. These items are usually retrieved by a teacher for one of two purposes: 1) to examine the bank, or 2) to create a test. Once items have been retrieved the items are organized into a test according to certain characteristics. Please rate the importance to your subject area of the following systemization features.

This part of systemization is concerned with the way items are selected from the bank for review or for a test.

	Very Important	Important	Somewhat Important	Not Important	Don't Know
24) Select by topic (Unit)	___	___	___	___	___
25) Select by instructional objective	___	___	___	___	___
26) Select by cognitive level	___	___	___	___	___
27) Select by item type (e.g. multiple choice, short answer, etc.)	___	___	___	___	___
28) Select by difficulty level	___	___	___	___	___
29) Select by discrimination index	___	___	___	___	___
30) Select by unique identification number (I.D. number)	___	___	___	___	___

	Very Important	Important	Somewhat Important	Not Important	Don't Know
31) Select by visual inspection of items on computer screen	___	___	___	___	___
32) Select items randomly	___	___	___	___	___
33) Select by keywords in stem	___	___	___	___	___
34) Select by keywords in alternative responses	___	___	___	___	___
35) Select by diagram used in stem	___	___	___	___	___
36) Select by diagram used in alternative responses	___	___	___	___	___
37) Select by reference to book or book chapter	___	___	___	___	___
38) Select all items in the entire bank for printing	___	___	___	___	___

This part of systemization is concerned with more general features available after selection of items has been made.

	Very Important	Important	Somewhat Important	Not Important	Don't Know
39) Easy editing of item text	___	___	___	___	___
40) Easy editing of item graphic	___	___	___	___	___
41) Select a previously used test.	___	___	___	___	___
42) Display a copy of the test before printing	___	___	___	___	___
43) Provision for special directions to be included with different test sections	___	___	___	___	___
44) Generate a set of tests, where different items are used in each of the tests generated but based on same criteria	___	___	___	___	___
45) Generate a set of tests where the same items are used but the alternatives are scrambled	___	___	___	___	___
46) Generate a set of tests where the same items are used but the order of items is scrambled	___	___	___	___	___
47) Keep together items of the same type (e.g. all multiple choice items together)	___	___	___	___	___
48) Keep together items on the same topic	___	___	___	___	___
49) Ability to reorder items before printing	___	___	___	___	___
50) Ability to have special headers & footers printed	___	___	___	___	___
51) Ability to modify page setup, line spacing and margins	___	___	___	___	___
52) A variety of formats for numbering questions	___	___	___	___	___
53) Print answer keys for all versions of the test	___	___	___	___	___
54) Ability to merge together different test item banks	___	___	___	___	___

Automation:

Computers have the capacity to automatically provide many features to teachers who normally construct tests manually. Please rate the importance to your subject area of the following automation features.

	Very Important	Important	Somewhat Important	Not Important	Don't Know
55) A unique identification number should be automatically assigned to each item when it is placed in the item bank	___	___	___	___	___
56) A record of previous versions of each item should be maintained by the computer (modification history)	___	___	___	___	___
57) Date of last revision for each item	___	___	___	___	___
58) Number of times each item has been used in other tests	___	___	___	___	___
59) Last date item was used on a test	___	___	___	___	___
60) Security access by password	___	___	___	___	___

Assuming the tests can also be machine scored, which information should be maintained

	Very Important	Important	Somewhat Important	Not Important	Don't Know
61) Difficulty level	___	___	___	___	___
62) Discrimination index	___	___	___	___	___
63) Distractor information	___	___	___	___	___
64) Means for high, median and low group	___	___	___	___	___
65) Number of times response to each item was omitted	___	___	___	___	___

Other information you feel is important:

____

____

____

____

____

____

____

Appendix C

Teacher Comments

Following are the written comments made by teachers at the end of the survey.

1. I basically want to be able to store test items, modify them, and be able to easily rescramble them. I would ideally like to be able to give the same exam to the same class and have them not know it.

2. I would like to be able to construct a test (say a short answer, fill in the blank type) where each student has some set of questions requiring the same solutions but each student has a different set of numbers. Therefore cheating by copying would be totally impossible.

3. For Biology, and probably other subject areas, a testbank must be able to include provisions for inclusion of diagrams, charts, tables, and graphs. There must be a capability to "draw" directly or integrate from some other source or sources.

4. The students need to respond to questions in some kind of written form (they may incorporate graphs or diagrams). The questions need to generate answers from the students and make them apply their knowledge. Consequently, the limitations of multiple choice, true/false questions.

5. Any usefully developed item bank should be such that versions are available for the computers currently in use in the schools, that is IBM, Apple II or Apple IIe etc. Banks should be on diskettes which can be used on microcomputer systems. It should be possible for copies to be made so that more than one copy is available in a large department.

6. Security is very important. A readily usable item bank would save hours of work each term. Should be adapted for at least the Apple and IBM and if Apple, for both the II+ and IIe. Directions for printing on a variety of printers would be great.

7. Excellent idea - looking forward to seeing your presentation in June. How about being able to include standard forms (as in accounting)? I teach such a wide variety of subjects (Law 20; Typing 10; Math 15, 25; Reading 10; Comp. Pro. 10; etc.) that I have answered in a very broad fashion.

8. Availability of math symbols such as powers, radicals etc. Machine scoring and item analysis is essential for item bank to be of use to smaller jurisdictions.

9. The ability to key in different styles of multiple choice questions is vital. Including the various math symbols as part of the font set available is key to High School needs.

10. Simple (short) identification of each item is important for printing of quiz/test. Machine scoring/analysis (not too sophisticate) would really be welcomed.

11. Number of times responses to each alternative was made. If the computer could check if the same question was already in the bank.

12. A problem with the test item bank now in use is screen editing. This feature is essential. Selection is also a very important factor as there must be large selection of test item questions.

13 Since there is not to my knowledge a test item bank for English, such a proposal has very meritable considerations. Test banks should be able to merge multiple choice questions and written response questions so as to have both types of items on a test as opposed to strictly multiple choice or written response tests.

14. We have a Chemistry 30 testbank on the Mac. The major problem is that the input of questions is done by people who are not familiar with the subject area. There are many mistakes not just typographical but regarding possible answers etc. More time should be spent in editing testbanks.

15. I would like to stress the importance of being able to include and/or insert graphics into test items stored. Ability to add items to the bank (and delete).

16. A well thought out questionnaire! I found reading it quite technical, and my response may be slightly off base. I can't help but conclude that a computerized test item bank, would be a very useful support to the teacher, a great time saver, that will allow teachers to use time more effectively. Thanks and good luck with the study!

17. I'm not sure of the importance of some of these items because of my unfamiliarity with computer generated programs.

18. Having access to programs to facilitate testing is great. Major problem is input of the data by non-typists etc. It would be useful to have a hard copy of all the questions.

19. We are into Macintosh Plus! Basically I would like a software program for the Mac for Math that can do the following.
   a. I could interchange multiple choice to short answer.
   b. Hand pick questions with ease and questions can be randomized.
   c. Multiple choice answers can be randomized.
   d. Store questions and diagrams together

We have Microtest II, but can you imagine it does not have super/subscript.