TESTING THE RELIABILITY AND VALIDITY OF OCCUPATIONAL ASSESSMENT INSTRUMENTS DEVELOPED THROUGH RANDOM SELECTION OF PREVIOUSLY VALIDATED TEST ITEMS

By

Steven Craig Clark

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Teacher Education

1987

ABSTRACT

TESTING THE RELIABILITY AND VALIDITY OF OCCUPATIONAL ASSESSMENT INSTRUMENTS DEVELOPED THROUGH RANDOM SELECTION OF PREVIOUSLY VALIDATED TEST ITEMS

By

Steven Craig Clark

The purpose of this research was to calculate estimates of reliability and to gain measures of validity on selected examinations developed using the vocational/occupational test-item bank at Michigan State University. The writer investigated the possibility of developing reliable and valid assessment instruments through random selection of test items from occupationally specific item subbanks.

Using the Michigan State University vocational/occupational item bank, certified vocational/occupational education instructors from the fields of machine trades and mechanical drafting developed the tests using a test blueprint. Two hundred eighty-eight secondary vocational students were administered the test using test/retest methods. A randomly selected group of students was given the same form of the test on both occasions; the remaining students were administered a parallel form as a retest. The first sets of data were used for estimating test stability; the second sets were used for estimating test stability and equivalence.

Instructors were asked to rank (pass/fail) their students' knowledge at both the occupational level (across all subtest domains) and the domain level (individual subtest). These data were used as the criterion measure and were correlated with the actual test results to establish the concurrent validity of the test.

The Pearson product-moment coefficients of stability were statistically significant at the .01 level for both occupational exams. The Pearson product-moment coefficients of stability and equivalence were also statistically significant for both instruments. Tetrachoric correlations of stability for the combined domain (duty) subscore ranking were statistically significant, as were the tetrachoric correlations for subscore stability and equivalence.

Correlations between students' actual percentile rankings and instructor-provided rankings were statistically significant for machine trades. For mechanical drafting, the correlations were not significant. At the subtest (domain/duty) level, the results for both samples were statistically significant. However, based on the small sample size, the results lacked practical significance. The overall results indicated weak to moderate agreement between the two variables tested. The evidence led the researcher to conclude that the tests had little concurrent validity with regard to the instructors' rankings.

ACKNOWLEDGMENTS

For guidance, encouragement, and assistance, a special expression of appreciation is extended to Dr. George W. Ferns. As major professor, his advice and knowledge proved invaluable in the completion of this degree.

In addition, profound appreciation is extended to Dr. Irvin J. Lehmann, dissertation chairman, whose knowledge of educational and psychological research methodologies provided the expertise necessary to complete this work. I also wish to acknowledge the contribution of my committee members: Dr. Rex E. Ray, whose insightful comments helped me maintain perspective on the practical importance of the study; Dr. Charles A. Blackman, whose availability and academic objectivity I will always admire; and Dr. Samuel A. Moore, whose wisdom and professionalism served as constant reminders of my goal.

Sincere appreciation is expressed to Susan Cooley, who aided me as editor, typist, and adviser in preparing the dissertation manuscript.

Dr. Chris M. Olson, whose reputation in the field of vocational/occupational education and commitment to instructors in Michigan, opened the doors for me to conduct this research.

I am deeply indebted to my wife, Susan, for her patience and understanding, without which this research could not have been undertaken. I also acknowledge the support of my children, Lauren and Phillip, who have grown up in academe without complaining about the many sacrifices due to their father's time on task. Special thanks to my parents, here and on the West Coast, for their support and encouragement throughout my educational career.

TABLE OF CONTENTS

		Page
LIST OF TABLES		x
LIST OF FIGURES		xii
Chapter		
I.	INTRODUCTION TO THE STUDY	1
	Introduction	1
	Background of the Problem	2
	Statement of the Problem	4
	Purpose of the Study	5
	Importance of the Study	6
	Research Questions	8
	Research Methodology	9
	Assumptions	10
	Limitations of the Study	11
	Definition of Terms	12
	Organization of the Remainder of the Dissertation	18
II.	REVIEW OF RELATED LITERATURE AND RESEARCH	20
	Introduction	20
	Historical Overview of General and Vocational/ Occupational Education Evaluation	21
	Effects of Evaluation	26
	Tested Vocational/Occupational Education Evaluation Instruments	29
	Estimating Reliability and Validity of Criterion-Referenced Tests	34
	Reliability	34
	Validity of Criterion-Referenced Tests	37
	Test-Item Banking	40
	Summary	41
III.	RESEARCH METHODOLOGY	43
	Introduction	43
	Design of the Study	43

Page

Research Questions 43
Population and Sample 45
Inservice Sessions for Pilot-Test Instructors/
Administrators 47
Development of the Instruments 48
The Tests 48
The Instructor Ranking Sheets 49
Data Collection 55
Data Processing and Analysis 57

IV. RESULTS OF THE DATA ANALYSIS 62

Introduction 62
Results 63
Research Question 1 63
Research Question 2 65
Research Question 3 67
Research Question 4 69
Research Question 5 72
Research Question 6 75
Summary 79

V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR
FURTHER RESEARCH 81

Summary 81
Procedures 82
Findings and Conclusions 83
Machine Trades 83
Mechanical Drafting 87
Synthesis 90
Recommendations for Further Research 91
Epilogue 92

APPENDICES

A. VOCATIONAL PROGRAM EVALUATION MODEL 94

B. MECHANICAL DRAFTING AND MACHINE TRADES TASK LISTINGS
FROM THE V-TECS 95

C. SAMPLE MECHANICAL DRAFTING AND MACHINE TRADES
TEST BLUEPRINTS 102

D. CORRESPONDENCE 107

Page

E. SAMPLE INSTRUCTOR RANKING SHEETS FOR MECHANICAL DRAFTING AND MACHINE TRADES 112

F. DATA SUMMARY TABLES 117

REFERENCES 155

LIST OF TABLES

Table		Page
3.1	Duty Areas Selected for Testing and Number of Items Requested, by Pilot Site	50
3.2	Number of Items in Each Duty Area Available for Inclusion in Tests	53
4.1	Correlational Data for Machine Trades Test Form A2/ Retest Form A2	64
4.2	Correlational Data for Mechanical Drafting Test/ Retest, Form A2	64
4.3	Correlational Data, Machine Trades Test Form A2/ Retest Form B2	65
4.4	Correlational Data for Mechanical Drafting Test Form A2/Retest Form B2	66
4.5	Subscore Ranking Agreement and Tetrachoric Correlations: Machine Trades, Test/Retest Form A2	68
4.6	Subscore Ranking Agreement and Tetrachoric Correlations: Mechanical Drafting, Test/Retest Form A2 . .	68
4.7	Tetrachoric Correlations for Combined Subscore Rankings on the Vocational/Occupational Assessment Test	69
4.8	Subscore Ranking Agreement and Tetrachoric Correlations: Machine Trades, Test Form A2/Retest Form B2	70
4.9	Subscore Ranking Agreement and Tetrachoric Correlations: Mechanical Drafting, Test Form A2/ Retest Form B2	71
4.10	Tetrachoric Correlations for Combined Subscore Rankings on the Vocational/Occupational Assessment Test	71

Page

4.11 Students' Percentile Rank Achievement in Relation to Instructors' Rankings of Students: Machine Trades, All Sites Combined, in Frequency and Percentage of Total 72

4.12 Percentages and Chi-Square Results for Students' Rank Achievement in Relation to Instructors' Rankings of Students' Achievement: Machine Trades . 73

4.13 Students' Percentile Rank Achievement in Relation to Instructors' Rankings of Students: Mechanical Drafting, All Sites Combined, in Frequency and Percentage of Total 74

4.14 Percentages and Chi-Square Results for Students' Rank Achievement in Relation to Instructors' Rankings of Students' Achievement: Mechanical Drafting 75

4.15 Vocational/Occupational Assessment Test Subscore Averages in Relation to Instructors' Rankings: Machine Trades, All Sites 76

4.16 Instructors' Rankings in Relation to Students' Subscore Rankings: Percentage Agreement and Tetrachoric Correlation Data, Machine Trades 77

4.17 Vocational/Occupational Assessment Test Subscore Averages in Relation to Instructors' Rankings: Mechanical Drafting, All Sites 78

4.18 Instructors' Rankings in Relation to Students' Subscore Rankings: Percentage Agreement and Tetrachoric Correlation Data, Mechanical Drafting . 79

LIST OF FIGURES

Figure		Page
3.1	Research Design	44
3.2	Sample Data Table	59

CHAPTER I

INTRODUCTION TO THE STUDY

Introduction

Cast in the foundation of vocational/occupational education is the conviction that evaluation is linked to program improvement. For more than 69 years, the national vocational education acts and subsequent amendments have addressed with varying intensities the belief that some type of evaluation should be incorporated into the vocational education setting. The latest vocational education act, the Carl D. Perkins Vocational Education Act of 1984, P.L. 98-523, again emphasizes the need for evaluation. The thrust of the Perkins legislation concerns program improvement through a sound assessment program. The present legislation, if enforced, will mandate some local involvement in program evaluation.

Providing reliable and valid evaluation procedures for vocational/occupational education programs is a formidable task, one with which the state of Michigan has started to grapple. Michigan has included two types of evaluation procedures in its attempt to evaluate the vocational experience. Process and product evaluation measures are addressed in Michigan's yearly planning process for vocational/occupational education. Process evaluation is used to detect or predict defects in the procedural design of a program or course during the implementation or operational stage (Wentling, 1980). Michigan evaluates process by using the Program Review and Planning (PRP) model (Michigan Department of Education, Vocational-Technical Education Service, 1982).

Product evaluation is superficially addressed through the survey responses of students who were enrolled in a vocational/occupational program the previous year. Michigan's new Vocational Program Evaluation Model (MDE, V-TES, 1986) (see Appendix A) has developed into an evaluation system that places new emphasis on student achievement as a part of vocational/occupational education product evaluation. This prompted the development of criterion-referenced tests for selected vocational/occupational programs. Yet data concerning the reliability and validity of these tests are nonexistent.

Background of the Problem

The Perkins Act has made program evaluation a requirement. To remain within the guidelines for reimbursement, over a five-year period 20% of the state fiscal agents must employ some measure of program evaluation each year. This scheme was developed to try to obtain 100% participation in some type of program evaluation by 1990. Given these conditions, the Michigan Department of Education has adopted the following philosophy: "Vocational education leads to competence and competence is linked to performance. Vocational education must, at some point, be concerned with the assessment of performance" (Runkel, 1986). Consequently, Michigan, along with many other states, needs to develop a vocational/occupational student competency assessment process and appropriate instrumentation.

The evaluation requirements of the Perkins Act have caused considerable confusion and some controversy within vocational/occupational education. The requirement of student assessment in programs funded under this act is troublesome for many districts, both administratively and programmatically. The question of how to assess can be approached from different positions, making the "letter of the law" seem negotiable.

Assessment in vocational/occupational education may cause some discomfort to those versed in measurement of traditional general education subjects. The differences between vocational/occupational and general education are as notable as their similarities. The environment in which vocational/occupational students learn is unique to their specific program. That, in itself, tends to rule out a standardized examination. The notion of an occupationally specific standardized assessment tool falls short when one considers the vast array of regional and local employment needs within the same occupational area. The differences in equipment and personnel within the same occupational specialty must also be considered. The need to assess vocational/occupational students' competency with valid and reliable instruments that reflect local programming provided the impetus for this research.

Statement of the Problem

The availability of occupational assessment tools is limited. Some vocational/occupational education programs have no commercially produced evaluation instruments from which to select, whereas the focus of some nationally normed, commercially produced standardized tests is unsatisfactory (Seney, 1986). The dilemma of providing assessment instruments that meet the needs of local programs nationwide is substantial. Even more important is the question of local autonomy.

Occupational program evaluation can be approached in many ways; no single evaluation procedure can be followed in all cases (Wentling, 1980). As with any content area, one instrument cannot provide all programs with a satisfactory measure of student achievement because validity is specific rather than universal.

In an attempt to alleviate the aforementioned problems, the Michigan Department of Education contracted with Michigan State University to develop an item bank of vocational/occupational test questions that address a variety of duty areas within selected vocational/occupational specialties. Having such a test-item bank would enable local instructors of vocational education programs to custom design tests that address specific local needs, thus preserving Michigan's commitment to local control of educational programming. However, the reliability and validity of vocational/occupational tests constructed using the test-item-banking method developed by Michigan State University have not yet been estimated.

Purpose of the Study

Because no current data are available regarding estimates of reliability and validity of examinations developed using Michigan's vocational/occupational test-item bank, the researcher's purpose in this study was to calculate such estimates on selected tests constructed and administered during the pilot testing of Michigan's occupational student assessment program. The vocational/occupational areas selected for this study were machine trades and mechanical drafting. Using data collected during the pilot study conducted in spring 1986, the researcher investigated the assumption that random selection of test items by duty and task areas would produce valid and reliable tests for the product evaluation of vocational/occupational students. By determining estimates of reliability and validity on selected Michigan pilot tests, the researcher obtained a preliminary measure of two important characteristics (reliability and validity) of an assessment instrument.

The writer selected commonly accepted measures of test reliability and validity for use in this study. The two reliability estimates used were (a) stability and (b) stability and equivalence. In addition, one type of criterion-related validity was studied: concurrent validity. The procedures employed in the research are discussed in detail in Chapter III.

Importance of the Study

The need to evaluate vocational education students is no different from the need to evaluate students in other curriculum areas. The results of evaluation can be beneficial for the instructor of the vocational program, as well as for the student who participated in the evaluation process. Measurement and evaluation are essential to sound educational decision making (Mehrens & Lehmann, 1978). The Carl Perkins Vocational Education Act of 1984, P.L. 98-524, emphasizes the need for vocational program improvement and requires evaluation of both the program's process and product. The vocational education philosophy adopted in Michigan is that competence in vocational/occupational education must be linked to performance and, at some point, include an assessment of performance. For this reason, the Michigan Student Occupational Assessment model was developed for the product evaluation of Michigan's vocational students (MDE, V-TES, 1986).

The need for an assessment test became evident when the Michigan Department of Education, realizing that major concerns existed among Michigan vocational administrators concerning the assessment issue, reevaluated the requirements of the Perkins Act (five-year evaluation plan; 20% participation the first year, 40% the second, and so on, to be started in spring 1986). Personnel in the Department of Education's Vocational-Technical Education Service realized that a vast amount of research and development was needed for assessing the state's vocational/occupational students to comply with the Perkins Act. Therefore, the Act's requirements for participation in student assessment were relaxed to allow for an extra year of developmental work. This provision, however, forces 25% accumulative participation each year over a four-year period.

The vocational program evaluation model was designed so that local vocational program instructors can identify specific vocational education program duties and tasks, using a master list developed by the Vocational-Technical Education Consortium of States (V-TECS; see Appendix B) for their vocational/occupational specialty. The selected duties and tasks are then entered into a computer program containing an unlimited number of test items, which are classified by duty and task. The MSU vocational/occupational test-item bank was developed using previously validated test items from occupational assessment tests used in V-TECS member states. The local vocational instructor requests a particular number of multiple-choice test items within the selected duties and tasks, and the computer randomly selects the items.

This research has possible national significance because of V-TECS involvement. The 22 member states in V-TECS are interested in the results of the Michigan pilot study. Research and development activities in Michigan intensified when the Michigan Department of Education V-TES awarded Michigan State University additional funds to further develop the assessment model, based on the pilot-testing effort. The format used in "cross-walking" or coding duties and tasks for this research can be used in a state's occupational education program because the coding employs V-TECS lists. Thus, the item bank developed at Michigan State University theoretically could be used by any V-TECS member state.

In addition, the findings of this study should help local education agencies and intermediate school districts understand the role this method of evaluation has in predicting the level of vocational/occupational knowledge of students receiving training. Such knowledge should help with curriculum decisions. The study could be of value to the Michigan V-TES program. This research is the first investigation of Michigan's test-item-banking system for vocational/occupational student assessment. Results of the study may help in determining whether future use of objective exams in vocational/occupational education programs is justifiable, in view of their reliability and validity.

Limited information is available regarding test-item banks. However, it seems appropriate to mention that Orenstein (1981) found three major weaknesses in using test-item banks: (a) validity information was not available, (b) reliability estimates were unavailable, and (c) instructional materials that complement tests were not available.

Research Questions

Each of the following research questions was posed to investigate the relative strength of the relationship that existed between sets of variables.

> Research Question 1: How stable are test scores obtained on the vocational/occupational assessment test?

This question concerns the stability of test scores.

> Research Question 2: How stable and equivalent are scores obtained on parallel forms of the vocational/occupational assessment test?

This question concerns the stability and equivalence of test scores.

> Research Question 3: How stable are subscore rankings obtained on the vocational/occupational assessment test?

This question concerns the stability of subscores within the test. Such consistency is of paramount importance because the assessment instruments were constructed according to duties (domains).

> Research Question 4: How stable and equivalent are subscore rankings obtained on parallel forms of the vocational/occupational assessment test?

This question concerns the stability and equivalence of subscores within the test.

> Research Question 5: How valid are the percentile rankings obtained on the vocational/occupational assessment test?

This question concerns the concurrent validity of the test.

> Research Question 6: How valid are the subscore rankings obtained on the vocational/occupational assessment test?

This question concerns the concurrent validity of the subtest scores.

Research Methodology

To fulfill the purposes of the study, the first step was to conduct a thorough review of the literature concerning evaluation and testing of vocational/occupational education students. The review included not only the theories and models of evaluation for vocational/occupational education students, but a historical overview of educators' views of testing student achievement for program improvement (process evaluation) and evaluation of program outcomes (product evaluation). Also, commercial tests developed for the evaluation of vocational/occupational education programs were examined. Finally, criterion-referenced reliability-estimating procedures were reviewed and analyzed, along with methods of measuring content and concurrent validity as related to the research questions under investigation.

Based on the literature regarding the various ways to gain reliability estimates, the researcher chose to use test-retest methods. Measures of concurrent validity were obtained using the vocational/occupational instructors' ratings as criterion measures. Content validity was used to measure the validity of the tests.

Assumptions

The researcher assumed that the vocational instructor content specialists who participated in this research by validating and coding V-TECS test items to V-TECS duty and task lists were representative of vocational instructors in the field. He also assumed that the vocational instructors involved in the study would use local curriculum guides as well as outside resources (advisory committee members, local industry/business standards, and minimum performance objectives) in determining duty and task selections for construction of the test instrument.

The investigator assumed that the vocational/occupational instructors involved in the study, having worked closely with the examinees for not less than one school year, would possess the appropriate knowledge and insight to permit them to give valid ratings.

The investigator assumed that the students, classrooms, and instructors involved in this study were typical of their counterparts in the general population of vocational programs. In addition, he assumed that the students who participated in the study would answer the test items in both administrations to the best of their ability. The investigator assumed the tests were content valid, based on the manner in which they were constructed.

Limitations of the Study

The study was limited in the following ways.

1. The writer did not attempt to validate the use of a test-item bank for any area outside of the vocational/occupational education curriculum at the secondary level.

2. The study was limited to selected areas of vocational/occupational education that received state or federal government funding during the 1985-86 school year.

3. The study was limited to the specific vocational/occupational items entered into Michigan State University's mainframe computer by April 15, 1986.

4. The researcher did not attempt to evaluate programs (local, state, or national), program instructors, or students.

5. The investigation was limited to vocational education programs of machine trades and mechanical drafting, whose instructors

volunteered to become involved in the study. Hence, instructors were self-selecting.

Definition of Terms

Inasmuch as this study relates to a specific area of education--vocational/occupational education--it is appropriate that definitions be provided in case certain terms are unfamiliar to the reader. Definitions provided by the Michigan Department of Education and the Vocational-Technical Education Consortium of States are paraphrased to provide continuity of meaning. All terms are defined with regard to this research.

Affective test item. An item designed to measure attitudes or values.

Cluster program. A group of closely related occupational courses with a common core of skills.

Coding. The process of assigning alphanumeric descriptors to items so that they can be identified.

Cognitive test item. An item pertaining to knowledge (as defined in Bloom's taxonomy), which is usually evaluated by means of a written test.

Competency-based education (CBE). A systematic approach to instruction, with precisely stated student outcomes, carefully designed student-centered learning activities, established guidelines for time-on-task for mastery, and set criterion measures for moving from one task to another.

Concurrent validity. The relationship between the test scores or other measures (predictors) and some independent external measure (criterion) obtained at the same time.

Content validity. The degree to which the sample of test items represents the domain about which inferences are to be made.

Cross-walking. The process of coding a test item to a specific task.

Criterion-referenced tests. Tests whose results are interpreted in terms of whether students reach or do not reach an established criterion.

Distractors. The incorrect responses to a multiple-choice test item.

Domain-referenced tests. Tests that draw a random or stratified sample of items from a precisely defined content area or domain.

Duty. One of the distinct, major activities involved in performing a job; each duty comprises several related tasks.

Employability skills. The generic skills related to seeking, obtaining, keeping, and advancing in an occupation.

Enabling objective. An objective related to background knowledge or skills that are prerequisite to mastering a certain task.

Error variance. The total effect of the chance differences between persons that arise from factors associated with a particular measurement (e.g., the participant's mood the day of the test).

Task analysis. The process of identifying (a) the performance elements of a task, (b) the knowledge and skills required to perform the task, (c) the tools and equipment necessary for job performance, and (d) the safety factors associated with efficient task performance.

Item reliability. The correlation between the participants' responses to a particular item and their total test scores.

Item validity. The correlation between the participants' responses to a particular item and their total scores on the criterion measure.

Job analysis. The systematic collection and examination of job-relevant information, such as duties, tasks, tools, equipment, references, and demographic information.

Job incumbent. An individual who is currently employed and performing competently in a work role.

Local validation of duties. The process whereby local vocational instructors select duty areas from the V-TECS list that are included within their specific curriculum.

Local validation of tasks. The process whereby local vocational instructors modify the larger, first form of their test by eliminating test items that they believe do not relate directly to their program's content/curriculum.

Norm-referenced tests. Tests that relate an individual's scores to those of a well-defined group (the norm). The goal is to differentiate clearly among students at different levels of achievement. Such tests do not pinpoint participants' individual strengths and weaknesses, just their total score relative to that of the norm group.

Objective referenced tests. Tests that measure learner performance in a way described in a performance objective and evaluate the learner against the criteria stated in the objective.

Occupational competency. The degree to which a student displays skills/knowledge in an occupational specialty, based on established criteria.

Performance objective. A written statement that describes a task the student must perform.

Pilot testing. Trying out an instrument or process on a group that is smaller than but similar to the actual test sample to determine the usefulness of that instrument or process.

Psychomotor test item. An item that measures specific performance requiring a muscular or motor skill.

Standardized test. A test that is administered using the same conditions (directions, materials, scoring, timing, identical questions) for every person who will take the test. Such a test typically has norms.

Stem. The statement or question portion of a multiple-choice test item.

Student achievement recording process. The process by which a student's achievement is documented so that tasks in which the student demonstrated competence can be recorded.

Student certification. The awarding of diplomas or documents certifying competency; or the advanced placement in a course sequence on the basis of test performance.

<u>Subject-matter expert</u>. A highly experienced job incumbent who assists in identifying task-performance steps and the skills and knowledge necessary to perform the job competently.

<u>Summative program evaluation</u>. A method used to determine a program's overall merit relative to its competition.

<u>Task</u>. A discrete unit of work that is performed by an individual, which has a definite and observable beginning and end within a limited time.

<u>Test administrator</u>. A person formally charged with giving a test to a group of participants.

<u>Test Form A1</u>. The preliminary form of the vocational student assessment test. It usually contains 40% to 50% more items than requested, to ensure sufficient items for the local validation process.

<u>Test Form A2</u>. The modified version of Test Form A1. After local validation of test items, this is the final form of the test.

<u>Test Form B1</u>. The second form of the vocational student assessment test, hypothesized to be an equivalent form of Test Form A1. It normally contains 40% to 50% more items that requested, to ensure sufficient items for the local validation process.

<u>Test Form B2</u>. The modified version of Test Form B1. After local validation of test items, this is the final form of the test.

<u>Test-item bank</u>. A collection of criterion- or domain-referenced test items coded for input and retrieval by computer or manually.

Test reliability. The level of internal consistency or stability of the measuring device over time.

Test stability. The ability of a test to yield consistent results over a period of time, given the same test and the same participants.

Test validity. The degree to which a test measures what it purports to measure.

Tetrachoric correlation. A correlational statistic used when both variables being correlated are in the form of artificial dichotomies.

Vocational education. Programs designed to prepare individuals for employment, or additional preparation for a related career requiring other than a baccalaureate or an advanced degree.

Vocational process evaluation. A type of evaluation whose primary goal is to detect or predict defects in the procedural design of a program or course during the implementation or operational stage (Wentling, 1980).

Vocational product evaluation. A type of evaluation that attempts to examine the outcomes of the vocational education effort.

Vocational program. A program of study specifically designed to prepare individuals for employment in a specific occupation or cluster of closely related occupations or to prepare them for a career requiring further education.

V-TECS. An acronym for the Vocational-Technical Education Consortium of States.

V-TECS catalog. Contains a description of the duties, tasks, performance objectives, performance guides, and related data for a specified occupation.

Organization of the Remainder of the Dissertation

Chapter II contains a review of the literature related to this research. A wealth of material exists on the effects of evaluation on both general and vocational/occupational education programs. Philosophical concerns are reviewed because the results of this study should have major significance for curriculum, guidance, and educational decision making. The second section of Chapter II includes a review of research and literature related to determining the reliability and validity of criterion-/domain-referenced tests.

In Chapter III, the writer chronicles the procedures followed in the study, from the development of the test-item bank through the inservice activities conducted with pilot-site instructors and administrators. The population is described, and the rationale for its selection is explained. The data-collection procedures are reported, including testing procedures and random selection of the retest groups. Data processing is then described in terms of item analysis, the subdivision of tests into domain-specific units for analysis, and comparative statistics that were used to estimate reliability and validity.

Chapter IV contains the estimates of reliability and validity. The findings of the pilot test are displayed in tabular format.

Chapter V contains a review of findings, conclusions drawn from the findings, recommendations, and suggestions for additional research.

CHAPTER II

REVIEW OF RELATED LITERATURE AND RESEARCH

Introduction

Numerous reports have been published on field-tested evaluation measures for both criterion-reference and norm-referenced vocational/occupational education instruments. Most of these reports concerned standardized measures, unlike the Michigan model, which uses a process of test construction based on local programming. The researcher learned that no research had been conducted on the estimations of reliability or measurements of validity on any assessment instruments developed and administered through the Michigan vocational program evaluation model.

To provide a background, perspective, and rationale for this study, the review of related literature and research is focused on four topics:

1. A historical overview of both general and vocational/occupational education evaluation, with emphasis on the latter, addressing program evaluation as well as student achievement.

2. A discussion of the effects of evaluation.

3. An examination of the available vocational/occupational evaluation instruments that have undergone some degree of controlled field testing.

4. A review of pertinent literature related to securing reliability and validity estimates for criterion-referenced tests. Also discussed is the limited research that has been conducted in the area of test-item banking and its relationship to the present investigation.

Historical Overview of General and Vocational/ Occupational Education Evaluation

Competitive examinations date back to 220 B.C. The Chinese used elaborate methods of evaluation in their civil service testing system (Wentling, 1980). Evaluation in occupational education may well have started in the early days of Greek sword and shield production, around 447 B.C. From Will Durant's (1939) comment that "the Spartan mother's farewell to her soldier son [was] 'Return with your shield or on it'" (p. 81), one gets the feeling that the first form of product evaluation was survival. In the early 1900s, Thorndike established measurement techniques to assess changes in human behavior. This was the first major report of content goals and objectives being used in evaluating human subjects (Thorndike & Hagen, 1969).

Later, in the 1930s, Tyler directed the Eight-Year Study, which emphasized the evaluation of student outcomes, specifically using college entrance exams for one group of students and matching them with students having similar backgrounds who were exempted from taking the exams to enter college (Chamberlin, Chamberlin, Drought, & Scott, 1942). In the Coleman Study of the late 1950s, the vocational opportunities for minorities in the United States were evaluated (Wentling, 1980). Such studies examined students' behavior, as well as outcomes of the educational experience.

In the 1960s, a study paralleling Coleman's work was completed by John Flanagan of the American Institute for Research. He and his associates evaluated 440,000 subjects in school settings and then attempted to associate this evaluation of abilities with job success or failure. This work was titled Project TALENT (Flanagan et al., 1964). Later, in 1969, Taylor initiated the National Assessment Project to assess school-to-school and state-to-state data, using student performance data as the vehicle for assessment.

The aforementioned projects were federally funded or supported by grants from private foundations. The evaluation of the effect of general education on American youths made good press and gained national attention. However, evaluation of vocational education programs in the United States was largely ignored, partially because of the separation of these programs from general education. The evaluation effort in vocational education came about primarily through stimulation provided by federal legislation. With the passage of the Vocational Education Act of 1963 (P.L. 88-210), each state became responsible for evaluating its programs through the use of state advisory committees (Wentling, 1980). In theory, this was to have been the first real attempt to evaluate the vocational effort. In practice, however, this maiden attempt fell short of its goal.

Two years after the Vocational Education Act of 1963 was passed, another piece of federal legislation was adopted, which forced major strides in educational evaluation. Passage of the Elementary and Secondary Education Act of 1965 (P.L. 89-10) "may represent the most important change in national domestic policy since the New Deal period" (Mernato, 1965). "The ESEA legislation required that each project conducted under Titles I and III possess a specific evaluation component to include a plan for the evaluation of process and product" (Wentling, 1980).

The Vocational Education Amendments of 1968 (P.L. 95-76) to the 1963 Vocational Education Act of 1963 re-emphasized the great need for program evaluation by the state advisory councils. The act referred to evaluation as follows:

1. The National Council shall review the administration and operation of vocational education programs, including the effectiveness of such programs in meeting the purposes for which they were established.

2. The National Council shall conduct independent evaluations of programs carried out under the 1968 amendments.

3. The National Council shall review the possible duplication of programs at the post-secondary and adult levels within geographic areas.

4. The State Advisory Council shall evaluate vocational education programs, services, and activities assisted under this act.

In addition, the Vocational Education Amendments set standards for third-party evaluation teams. These teams reported to the project staff and the funding agency, part of the checks-and-balances system established by the new amendments based on program evaluation being directly tied to the funding of programs (Evans & Herr, 1978).

The 1976 Education Amendments to the 1963 Vocational Education Act again emphasized and expanded the mandates for vocational education to involve evaluation as part of its ongoing plan (P.L. 94-482, 1976). The Amendments specifically referred to 28 different forms of state and local evaluation under Title II of the Amendments, with dates for the completion of specific evaluation activities (Wentling, 1980).

Specifically stated, the rules and regulations read:

> The state board shall, during the five-year period of the state plan, evaluate in quantitative terms the effectiveness of each formally organized program or project supported by federal, state, and local funds. These evaluations shall be in terms of:
>
> Results of student achievement as measured, for example, by:
> - Standard occupational proficiency measures
> - Criterion-referenced tests
> - Other examination of students' skills, knowledge, attitudes, and readiness for entering employment successfully. (P.L. 94-482, sec. 104.402)

In the late 1970s and early 1980s, evaluation took on new meaning. The late 1970s marked the beginning of the "effective-schools" movement (Brookover, 1979) and the "effective-schools" research (Edmonds & Lezotte, 1979, 1980), both of which identified "correlates" (Edmonds, 1979) that were felt to have a profound influence on the effectiveness of schools. One of these correlates was "careful monitoring of student achievement as a basis for program evaluation" (Edmonds, 1979). Again, there is strong evidence for the need to assess students, regardless of the curriculum in which they are involved, "for sound educational decision making" (Mehrens & Lehmann, 1978).

In April 1983, the National Commission on Excellence in Education published a report entitled A Nation at Risk: The Imperative for Educational Reform (National Commission of the States, 1983). In this report, the term "test" is used 11 times in reference to factors that have contributed to the "risk" (Independent School, 1983).

"The excellence movement has given scant attention to the purpose of secondary and postsecondary vocational education" (Bottoms, 1984). This statement by the director of the American Vocational Association (AVA) expresses the feelings of most vocational educators after the reports on excellence in education and their subsequent recommendations were released. The Michigan Council of Vocational Administrators (MCVA) concurred with the AVA director's report and, in addition, stated that "a concentration on the need to analyze, evaluate, and improve our current performance in delivering the basics in education to our youth is long overdue" (Pratt, 1984). Evaluation remains one of the most important factors in the educational process.

With the volumes of educational reports and recommendations that emerged in the early 1980s reporting the benefits of student program evaluation, the Carl Perkins Vocational Education Act of 1984 (P.L. 98-524, 1984) paralleled the national attitude--that student/program evaluation leads to improved programs. The Perkins Act requires both process and product evaluation of programs at the local level; 20% percent of the fiscal agents must evaluate students by the end of the 1985-86 academic year. The vocational education evaluation policy adopted in Michigan states that "vocational education leads to competence and competence is linked to performance. Vocational Education must, at some point, be concerned with the assessment of performance" (Runkel, 1986). These words from Michigan's top public school administrator parallel those of the Perkins Act and respond to calls for accountability at all levels of education (Smith, 1985).

Effects of Evaluation

The National Institute of Education (1985) released survey results on the educational attitudes of the American people. The data showed that the American public strongly favored testing for promotion from junior high school, for graduation from high school, and for determining teacher competency. However, teachers are not to neglect their students' social and personal development while they emphasize academics (National Institute of Education, 1985). The effects of evaluation on both the school as an institution and the students within schools encompass a wide range of philosophical and sociological ideologies. Both evaluation for planning and evaluation for learning are addressed in the present study, with the analysis of programs by instructors, and development of student assessment tools based on the instructors' local curriculum content. As Dressel (1954) stated,

> Evaluation incorrectly done is at odds with the promotion of learning. Evaluation correctly done should enhance learning because it aids both the teacher in teaching and the student in learning.

The student and the institution can be directly influenced by the evaluation process. The results of school achievement tests can alter people's attitudes toward education. Lezotte (1986) pointed out that an evaluation model that uses achievement testing to monitor student progress and gains in student achievement as measured on standardized tests has merit and is "getting a good reputation."

The subject of testing often conjures up notions of exclusion. However, "there is no good evidence that accurate feedback damages students' self-concepts, but there is much evidence that such feedback improves subsequent performance" (Mehrens & Lehmann, 1978). Students can even learn during the testing process. "It is probably not extravagant to say that the contribution made to a student's store of knowledge by the taking of an examination is as great, minute for minute, as any other enterprise he engages in" (Stroud, 1946).

The institution can benefit substantially from the evaluation process if the data are used as they were intended to be used--for sound educational decision making. Speaking to the General Subcommittee on Education of the House of Representatives' Committee on Education and Labor, Parnell (1973) stated:

> Measurement is the hand-maiden of instruction. Without measurement, there cannot be evaluation. Without evaluation, there cannot be feedback. Without feedback, there cannot be good knowledge of results. Without knowledge of results, there cannot be systematic improvement in learning.

The improvement of learning and instruction through evaluation of students and analysis of test results leads one to the assumption that assessment testing can solve all of education's woes. That assumption is largely unproven; however, many writers have supported the use of assessment testing in United States schools (Olson, 1986).

Gordon Cawelti (1986), executive director of the Association for Supervision and Curriculum Development, spoke of the effective-schools movement and evaluation:

> The effective-schools movement has provided a very useful focus for large numbers of schools to concentrate energies in the direction

> of improving student achievement. It certainly has weaknesses, but it also has compelled school districts to clarify their academic goals and work on raising expectations for kids and monitoring progress.

General education curriculum in the schools should focus as much on society as vocational education curriculum does on the world of work. "Only rarely has curriculum content not reflected what is happening outside the school" (Apple, 1979). With "achievement" and "accountability" as the buzz-words in society today, the curriculum changes being implemented reflect those feelings (Naisbitt, 1984). Many educators believe that assessment for educational decision making at the building level is less than perfect. "Most curriculum decisions are made by the building principals with quite a bit of input from teachers; parents and students, however, are generally left out" (Kimpston & Anderson, 1982). This leads one to believe that test scores are not considered in the change process and that little parental input is taken into account.

In the late 1930s, many educational leaders adopted Frederick Taylor's "scientific management" theories. Taylor (1938) asserted, "The curriculum must be efficiently managed and outcomes precisely predicted. Ten years later, in Basic Principles of Curriculum and Instruction, Tyler (1949) asked four questions about the curriculum-development process:

1. What educational purpose should schools seek to achieve?
2. How can learning experiences be selected that are useful in attaining objectives?

3. How will learning experiences be organized for effective instruction?

4. How can the effectiveness be evaluated?

Tyler's model of curriculum development had evaluation as a major component.

From the late 1940s until Sputnik, educational curriculum reflected the economic stability of the times. Stufflebeam (1971) developed one of the more generalizable curriculum and evaluation models, which has become popular today. His model includes four types of evaluation: context, input, process, and product. Ten years later, Stufflebeam (1981) chaired the Joint Committee on Standards for Educational Evaluation, whose model for educational-program evaluation comprised 32 standards categorized under four major headings: utility, feasibility, propriety, and accuracy. The last category was research based and dealt primarily with evaluation.

At present, evaluation and assessment are keys to federal- and state-funded programs. Funding is tied directly to student and program (curriculum) evaluation (P.L. 98-524).

Tested Vocational/Occupational Education Evaluation Instruments

This section of the literature review includes an overview of field-tested measures of skills and knowledge that are occupationally specific (e.g., machine trades). The tests included in this section encompass the field-tested measures compiled through computer searches of data bases such as the Educational Resources Information Center (ERIC), the Vocational and Occupational Information Center for

Educators (VOICE), Vocational Education Curriculum Materials (VECM), and Dissertation Abstracts. The writer searched these sources to obtain information on instruments that had been developed or research that had been conducted within the past ten years. Other sources of information on occupational/vocational education testing were acquired through the following test-development institutions: the National Occupational Competency Testing Institute (NOCTI), the American Institute for Research (AIR), and the Instructional Materials Laboratory of The Ohio State University. The vast majority of the verified, field-tested instruments were developed at these three institutions (Lewis & Martin, 1986).

Through the National Center for Research in Vocational Education, vocational/occupational researchers Lewis and Martin (1986) under contract grant from the United States Department of Education, compiled a compendium to help vocational educators select appropriate test instruments to measure the learning outcomes of their programs. The compendium comprised 1,200 entries found in the Vocational Education Curriculum Materials (VECM) data base. Only 150 validated instruments were included in the compendium. This section of the review is focused on those instruments.

Half of the 150 verified, field-tested evaluation instruments were developed at the National Occupational Competency Testing Institute (NOCTI). Of that number, 50 are teacher occupational competency tests, designed for experienced workers who wish to apply for teaching positions in vocational education programs. The remaining tests developed through the NOCTI--the Student Occupational Competency Achievement Tests--are specifically designed to measure students' occupational knowledge and skills.

The NOCTI exams are standardized, norm-referenced instruments developed by the Institute. Users of these tests must ensure the security of the instruments and return all testing materials. Users purchase a testing service, not actual test materials (Lewis & Martin, 1986).

The AIR tests include both written and performance sections, as do the NOCTI exams. The AIR tests also include a Work Habits Inventory, a three-questionnaire packet containing the same 49 rating scales; the student completes two of the questionnaires, and the instructor completes the remaining one. The students' first questionnaire concerns "how you think the average employer wants employees to behave," and the second refers to "the extent to which each of the descriptions applies to you." The teachers' questionnaire is designed to ascertain the degree to which the 45 descriptions apply to the student. General employers' ratings are included, to compare ratings across all questionnaires, as recommended by the test developers.

This questionnaire approach is relevant to the present research in regard to the criterion-related validity estimates the researcher made for the Michigan tests developed and used in this study. Subjective judgments, gathered through questionnaires completed by both students and instructors, are valuable tools in assessing specific measures of validity (Bloom, Hastings, & Madaus, 1971). Bloom et al. commented on these subjective judgments: "It is true that careful subjective judgements by those [instructors] who have experience with performers may substitute as estimates until empirical evidence is collected" (p. 69).

The Ohio tests are written tests designed to measure students' occupational knowledge and skills. Administered as part of the Ohio tests is the California Short Form Test of Academic Aptitude, which takes about one hour to complete. The developer recommends that the latter test be administered so that the students' achievement percentiles can be compared to their aptitude percentiles. This comparison is possible because both the Ohio tests and the California Short Form Test of Academic Aptitude were normed on the same population.

Like the tests developed by the NOCTI, the Ohio tests are standardized and norm referenced. Users of the Ohio tests must ensure the security of the tests and return all test materials. A testing service is purchased, rather than the actual test materials.

The Ohio tests are meaningful to this research because they involve comparison of student percentile rank, as determined by the instruments, with a criterion measure of percentile rank, assessed by another instrument, both normed on the same population. The measured percentile rank and predicted percentile rank of the test population used in the present research were compared using much the same methodology.

The aforementioned occupational/vocational education test-developing institutions gain estimates of test reliability and validity using traditional methodologies. However, the test developers considered in this section did not create their tests in the same manner in which the Michigan instruments used for this study were developed. Although many of the national tests (NOCTI, AIR, Ohio tests, and so on) can be considered criterion referenced in specific settings, they are by design occupationally specific, standardized, norm-referenced exams. The researcher would be remiss in eliminating them from this review; however, the direct relationship to the current research was tangential in nature.

One occupationally specific test developer uses a data bank of test items. The Interstate Distributive Education Curriculum Consortium (IDECC) develops tests by individual competency or by occupation and level. The IDECC item bank has a "descriptive test key," which provides a rationale and reference for each multiple-choice test item. The existence of the IDECC item bank is the major reason the Michigan item bank does not include any marketing-education test items. However, as testing of reliability and validity were of interest, this researcher contacted the director of IDECC testing and became keenly aware that test reliability and validity are not key concerns in the IDECC testing process (Gleason, 1986).

Pollock and McDone (1974) developed a national item bank for tests of driving knowledge. Materials intended for operational licensing and educational agencies to use in developing driving-knowledge tests were prepared. Multiple-choice-item pools were generated for four classes of certification; approximately 2,200 items were developed. Field tests were conducted to collect psychometric, normative, and validation data for one of the certification classes (Class C, passenger car and light truck); however, no reliability estimates were given.

O'Reilly (1976) summarized the development and validation of two different types of multiple-choice instruments. Thirty-six forms of the tests, which covered literal comprehension, were given to 5,722 students in grades 1 through 9. Upon examining the data, O'Reilly discovered that both forms of the test were "generally" consistent. He also estimated test reliability using the Kuder-Richardson 20 statistical formula. The reliability estimate for these exams was considered "very good."

Estimating Reliability and Validity of Criterion-Referenced Tests

Reliability

Of primary interest in this study were measures of relationships. If research produces two sets of scores from the same group of subjects, it is often desirable to know the degree to which those scores are related (Mehrens & Lehmann, 1978). Fox (1969) defined test reliability as follows:

> Reliability . . . is the basic attribute which every procedure must possess. . . . By reliability we mean the accuracy of the data in the sense of their stability, repeatability, or precision. A perfectly reliable data-collection instrument is one which, if administered twice under the same circumstances, would provide identical data. (pp. 352-53)

The researcher was also interested in criterion-referenced interpretation. Determining reliability is as important in studies of criterion-referenced tests as it is in research dealing with norm-referenced tests (Martuza, 1977). What is needed, however, because of the domains within tests, are techniques that yield meaningful, interpretable results, regardless of the amount of variability present in the test-score distribution (Martuza, 1977).

Several authors (Faggon, 1977; Mehrens & Lehmann, 1978; Popham & Husek, 1969; Sax, 1974) have agreed that determining the reliability of criterion-referenced measures can be problematic. Ebel (1979), however, did not concur. He stated,

> Reliability is as important for criterion-referenced tests as it is for conventional norm-referenced tests, and it can be determined in much the same way. The notion that "classical" test theory and "traditional" methods of test analysis are inappropriate for criterion-referenced tests is based on a misconception, namely, that scores on a criterion-referenced test show no variability because all who take the test make perfect scores, answering all the questions correctly. While this is a theoretical possibility, it almost never happens. One could "rig" a test to make it happen, but such a test would serve no useful educational purpose. The only reason for giving a criterion-referenced test is to identify students who have not achieved a certain educational objective. . . . The same procedures that are used to estimate the reliabilities of norm-referenced tests--that is test-retest, equivalent forms, split-halves, or Kuder-Richardson--can also be used to estimate the reliabilities of criterion-referenced tests. (pp. 281-82)

Although it is not the researcher's intention to solve the controversy surrounding this issue, it is important that there be as much confidence as possible in the data and findings of any research.

Fox (1969) explained that:

> To estimate reliability through the alternate-form procedure, the researcher must develop two parallel or equivalent forms of his instrument, administer both to the same people, and correlate the two sets of data obtained. The administration of the two forms can be with an intervening time interval. (p. 355)

Ebel (1979) commented in this regard:

> If the same test is given twice to the same group, if two different but equivalent forms are given to the same group, or if a single test is split into two equivalent halves that are separately scored, we use an index of the degree of agreement between the pairs of scores for each person as the basis for estimating the test reliability. (pp. 218-19)

Popham and Husek (1969) stated that although the issue of variability is at the core of the difference between norm-referenced and criterion-referenced tests, it is true that one almost always gets variant scores on any psychological test; but variability is not a necessary condition for a good criterion-referenced test. If the objectives within the test are substantially diverse, the items measuring them should be considered as different tests, not as a single, all-encompassing measure.

Humbleton and Novick (1973) further suggested: "An alternative measure of reliability might simply be the proportion of times that the same decision (pass/fail, master/nonmaster) would be made with the two parallel instruments" (p. 168). Carver (1970) agreed, but from a different perspective:

> The reliability of any test depends upon replicability, but replicability is not dependent upon test score variance. If examinees obtain similar (but not identical) scores on parallel forms of the same criterion-referenced test . . . near perfect replicability exists even though test reliability estimates using classical correlational methods (test-retest, parallel-forms, etc.) might be close to zero (no correlation). (p. 71)

Carver went on to explain how he believed criterion-referenced test reliability can be assessed:

1. The reliability of a single form of a criterion-referenced device could be estimated by administering it to two comparable groups. The percentage that met the criterion in one group

could be compared to the percentage that met the criterion in the other group. The more comparable the statistics, the more reliable the test.

2. The reliability of a criterion-referenced test could be assessed by comparing the percentage of examinees achieving the criterion on parallel tests. (p. 56)

Based on the preceding discussion, this researcher obtained estimates of reliability by employing three strategies suggested in the literature:

1. Test-retest reliability methodologies using the same sample.

2. Test-retest reliability methodologies using parallel forms of a test with the same sample.

3. Produced reliability estimates based on each subsection of the test, for both identical and parallel forms of the same test with the same sample.

Validity of Criterion-Referenced Tests

Unlike reliability, there is less controversy within the community of test theorists concerning acceptable methods for seeking estimates of test validity. Humbleton (1984) stated: "Criterion-related validity studies of criterion-referenced test scores are not different in procedure from studies conducted with norm-referenced tests" (p. 21).

Ebel (1979) described test validity as follows:

1. Validity is the most important quality a test can have.

2. Validity refers to the effectiveness of a test in measuring what it is intended to measure.

3. Validity has many different facets, takes many different forms, and can be defined in many different ways.

4. Validity can be demonstrated best by tryout after the test has been constructed.

5. To demonstrate validity, one must have criterion measures, real, synthetic, or hypothetical. (pp. 289-99)

Sax (1974) indicated that tests have content validity if "the behavior and subject matter called for in the items corresponds to the behavior and subject matter identified in the specific objective." Scriven (1977) described how content validity of an instrument might be obtained:

> [Get] some external judgments as to the cohesiveness of the alleged goals, the actual content, and the test question pool. Without this, the validity of the tests and/or the utility of the curriculum will suffer, possibly fatally. There is no need at all for the individual judge at this task to be a professional evaluator; indeed, professional evaluators are frequently extremely bad at this. A good logician, a historian of science, a professional in the subject matter field, an educational psychologist, or a curriculum expert are possible resource categories. (p. 349)

Predictive validity is one of two types of criterion-related validity. Sax (1974) wrote that:

> The term criterion-related validity . . . means the correlation of measurements within an external criterion. If the measurements are used to predict future behavior, these correlations are called predictive validity coefficients; if validity is estimated by correlations of measurements with currently obtainable criteria, these correlations are called concurrent validity coefficients.

In determining the predictive validity of an instrument, the researcher must have an external criterion or other indicator for the variable under study. Fox (1969) wrote:

> Concurrent, congruent, and predictive validity express validity in a correlation and so provide an estimate of extent. . . . All of these procedures to produce the statistical estimate of validity require the researcher to have available some other measure for the variable under study. (p. 371)

Bloom et al. (1971) described other measures that are considered viable in estimating the validity of a test as subjective judgments. Addressing knowledge acquisition, the authors noted that subjective judgments of those who have experience with the performers may substitute as estimates (validity) until empirical evidence is collected.

Popham and Husek (1969) discussed the validity of a criterion-referenced instrument as follows:

> Criterion-referenced measures are validated primarily in terms of the adequacy with which they represent the criterion. Therefore, content validity approaches are more suited for such tests. A carefully made judgement, based on the tests' apparent relevance to the behaviors legitimately inferable from those delimited by the criterion, is the general procedure for validating criterion-referenced measures. (p. 164)

Estimating reliability and validity of any test, either criterion referenced or norm referenced, depends on a number of factors, the most important being the student's score. Mehrens and Lehmann (1978) pointed out many of the problems inherent in testing students:

> A student's test score may vary for many reasons. The amount of the characteristic we are measuring may change across time (trait instability); the particular questions we ask in order to infer a person's knowledge could affect the score (sampling error); any change in directions, timing, or rapport with the test administrator could cause some variability (administrator error); inaccuracies in scoring a test paper will affect the scores (scoring error); and finally such things as health, motivation, degree of fatigue of the person, and good or bad luck in guessing could cause score variability. (p. 7)

Validity measures of the content- and criterion-related types, along with other validity measures, were discussed in the preceding section. With no empirical data existing on the Michigan vocational/occupational student assessment model, this researcher obtained measures of criterion-related validity using the subjective judgments provided by the pilot-site instructors as his external criterion. Content validity was measured during the local validation process (described in Chapter III) performed by the pilot-site instructors and through the V-TECS test-item writing process, done before items were inserted into the item bank.

Test-Item Banking

Test-item banking systems are having an important influence on educational test development. A few item banks exist today, and the benefits that can be achieved through item-bank test development have been extolled (Orenstein, 1981). Numerous articles have been written about particular item-bank operations (Eaddy, 1985; Faggen, 1978; McCage, 1986; Orenstein, 1981). However, despite the widespread proliferation of item banking in a number of different subject areas, definitive and comparative data regarding the use of item banks are lacking (Orenstein, 1981).

Orenstein discussed the strengths and limitations associated with the use of item banks for test construction. He described the strengths as follows:

> Time savings, cost savings, good item quality, test security, the construction of minimum competency tests, complete descriptions of content covered by items, schemes for classifying and coding items which provides for each item's retrieval, appropriate statistical information on item quality, and procedural operations for using the item bank.

Orenstein cited the absence of "validity information, reliability estimates, and instructional materials that complement tests" as limitations repeatedly voiced by test-item-bank users.

Test-item quality and validity have been a priority when developing items for use in item pools (Orenstein, 1981). The Vocational-Technical Education Consortium of States (V-TECS, 1985) developed The Handbook for Developing Criterion-Referenced Test Items to establish a standard by which member states could construct and validate test items specifically identified for use in item banks as well as in traditional vocational/occupational exams. The Handbook contains five sections that explain the process of writing and validating each item. The process includes such steps as assisting users in locating and verifying tasks, through task analysis, and field testing strategies for validation.

Summary

The present study is the first report on the student occupational assessment pilot study conducted at various sites in Michigan during spring 1986. Related concerns surrounding the present research--history of general education and vocational/occupational education evaluation and the effects of evaluation--were discussed in the first sections of the review. Over the years, educators have seen that measurement of student achievement is of major concern both within education (for program improvement/evaluation of student achievement) and outside the educational arena (for justification and advancement). The requirement of evaluation in vocational/occupational education by the Carl Perkins Vocational Education Act of 1984 (P.L. 98-524) has brought new attention to the process by which classroom evaluation is conducted.

Of the 1,200 occupational/vocational evaluation instruments mentioned in the literature, only about 150 have been field tested (Lewis, 1986). In addition to the Michigan State University occupational/vocational test-item bank, only two centers (IDECC and the driving-knowledge item bank) were discovered that develop occupational/vocational exams by accessing a data bank of multiple-choice test items. IDECC's method of developing exams parallels the procedure used to produce Michigan's exams. However, reliability and validity estimates for the IDECC exam were unavailable. The driving-knowledge item bank has been field tested, and both reliability and validity have been reported. However, the reliability estimates were gained using norm-referenced methodologies (Kuder-Richardson 20) as compared to test-retest methods, which were used in the present study.

The final section of the review dealt with the reliability and validity of criterion-referenced tests, as well as of some norm-referenced measures. Types of reliability and validity were discussed, along with arguments for and against specific methods of estimating reliability and validity. The value of this study was underscored, based on the importance testing "experts" have placed on estimating reliability and validity for sound assessment instruments.

CHAPTER III

RESEARCH METHODOLOGY

Introduction

Chapter III contains a discussion of the research methodology employed in the study, including the basic design of the study, the research questions, the population, pilot-test instructor inservice sessions, pilot-test instrument-development procedures, development and pilot testing of the instructor ranking sheets, data-collection techniques, and statistical-analysis procedures.

Design of the Study

The research design shown in Figure 3.1 is helpful in conceptualizing the steps involved in conducting this study. The various procedures followed in carrying out the investigation are explained in the following pages.

Research Questions

The study methodology was designed to test the research questions posed in this study. Those questions are as follows:

> Question 1: How stable are test scores obtained on the vocational/occupational assessment test?
>
> Question 2: How stable and equivalent are scores obtained on parallel forms of the vocational/occupational assessment test?

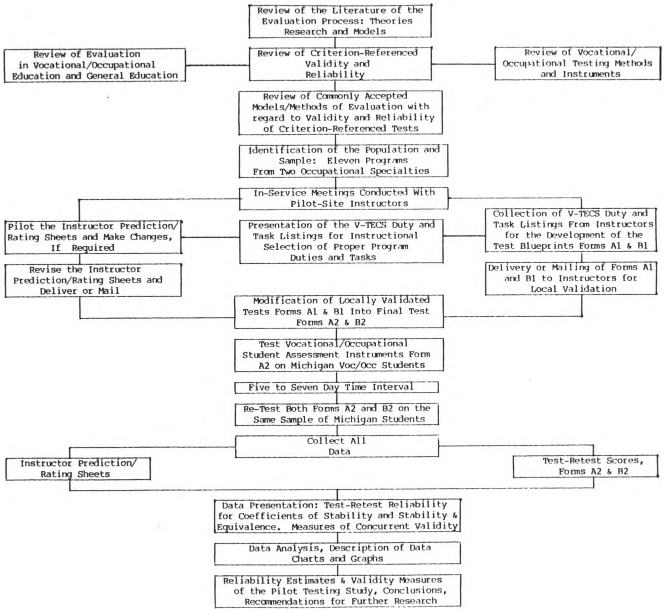


Figure 3.1: Research design.

Question 3: How stable are subscore rankings obtained on the vocational/occupational assessment test?

Question 4: How stable and equivalent are subscore rankings obtained on parallel forms of the vocational/occupational assessment test?

Question 5: How valid are the percentile rankings obtained on the vocational/occupational assessment test?

Question 6: How valid are the subscore rankings obtained on the vocational/occupational assessment test?

Population

In March 1986, Michigan State University personnel asked vocational/occupational education directors at both the secondary and post-secondary levels to nominate programs and instructors to pilot test a new assessment procedure (see Appendix D). The study comprised 11 occupational specialties (see Appendix D). Included in these specialties were 46 vocational/occupational education programs. (For simplicity, the term "program" as used in this dissertation refers to the specific site, instructor, and students enrolled in spring 1986.) Program inclusion was based on the following criteria: (a) the vocational/occupational education program received federal and/or state funding for the 1985-86 school year; (b) a certified vocational instructor was employed; (c) the program was an occupational specialty for which a test-item bank had been developed (a pool of items was available for 11 occupational specialties); (d) the program had administrator and instructor support for the pilot-testing effort, and these individuals volunteered to become involved in the study; and (e) the instructor had the flexibility in programming to allow for scheduling of both test and retest dates. Each administrator and instructor had an equal opportunity to participate in the vocational/occupational pilot testing conducted by the Michigan Department of Education V-TES.

From the vocational/occupational specialties meeting the aforementioned criteria, two occupational specialties--mechanical drafting and machine trades--were selected. All programs within these two specialties were included in the study. Four mechanical drafting programs and seven machines trades programs were identified, a total of 11 programs. The student enrollment in each program and the response rate (in parentheses) were as follows:

Mechanical Drafting		
Site H	n = 42	(39)
Site I	n = 25	(21)
Site J	n = 23	(18)
Site K	n = 27	(19)
	117	(97)

Machine Trades		
Site A	n = 11	(10)
Site B	n = 52	(37)
Site C	n = 31	(25)
Site D	n = 23	(21)
Site E	n = 20	(15)
Site F	n = 28	(22)
Site G	n = 6	(3)
	171	(133)

The total number of vocational/occupational education students involved in the test-retest procedure, and for whom participating instructors returned an instructor ranking sheet, was 230 or 80% participation.

Inservice Sessions for Pilot-Test Instructors/Administrators

During March and April 1986, the researcher met with each selected pilot-test-site instructor to demonstrate the process of creating assessment instruments using V-TECS duty and task listings. These meetings had two main objectives: First, the test-item bank was coded (cross-walked) to the V-TECS duty and task listings, not to any state curriculum guide. Therefore, exposure to and understanding the listings was important to the success of the testing effort. Second, the instructor ranking sheets (see Appendix E) used as criterion measures for validity were developed directly from the V-TECS duty and task listings. The ranking sheets were pilot tested at these meetings and at subsequent individual meetings between the researcher and program instructors.

The procedure for constructing the student assessment instruments was explained through the use of examples and scenarios. The researcher informed the pilot-test instructors about the process to be used to develop their student assessment tests. Those procedures were as follows:

1. Instructors chose from V-TECS duty and task listings for their occupational specialty those that constituted the content of their programs.

2. Participants indicated the number of items within the duty and/or specific task they wanted to appear on the test. The number of items per duty and/or task was based on the instructors' evaluation of the curriculum's emphasis on a particular duty or task area. The

maximum number of items generated on any assessment test was not to exceed 100.

3. The researcher developed test Forms A1/B1 and returned them to the instructors for the last steps of test development. Instructors validated all test items on the preliminary forms (A1/B1) and checked stems, distractors, and answers (from the answer key) for accuracy. Participants also identified items to be eliminated from the final version of the vocational/occupational assessment test (Forms A2/B2) for that instructor.

4. Instructors returned both forms of the assessment test to the researcher for modification (elimination of undesirable test items and/or addition of items) into the final version of the vocational/occupational assessment test (Forms A2/B2).

The instructors were also informed of the procedures they were to perform to ensure the quality and integrity of this research.

At the conclusion of the inservice meetings, test and retest dates were confirmed for each of the 11 programs involved in the study. Instructors were asked to inform their students of the pilot testing dates to ensure attendance.

Development of the Instruments

The Tests

Within a week of the inservice meetings, the pilot-site instructors returned their selections from the V-TECS duty and task listings to the researcher. Table 3.1 shows the duty areas selected for testing, along with the number of items requested from each duty area, for each pilot site. These data constituted the test blueprint upon which Forms A1 and B1 of the test were developed.

The tests for the two occupational specialties selected for this research (mechanical drafting and machines trades) were constructed at Michigan State University, using the occupational test-item bank that Michigan State University personnel had developed for the Department of Education. At the time of pilot testing, the number of items in each duty area available for inclusion in the tests was as shown in Table 3.2. There were at least three items for each task. It is evident that not all duty areas were well covered by test items at the time of pilot testing. The researcher assumed, however, based on the duty and task listings returned by pilot-site instructors, that items existed in the bank to address the research questions. Both forms of the tests were developed using items available in the item bank. The reader is reminded that the instructors controlled all items included in the final form of the test.

The Instructor Ranking Sheets

To collect data for studying concurrent validity, the researcher used instructor ranking sheets as a criterion measure (see Appendix E). Concurrent validity pertains to the empirical technique of studying the relationship between the test scores and some independent external measures (criteria) (Mehrens & Lehmann, 1978). The instructor ranking sheets were administered to the instructors at the same time as or within 10 days of the student retest. As stated in

Table 3.1.--Duty areas selected for testing and number of items requested, by pilot site.

Pilot Site	Duty Code & Description	# of Items Requested
	Mechanical Drafting	
H	F--Detailing Single Parts	25
	G--Drawing Layouts	20
	J--Drawing Layouts of Single Parts	20
	L--Drawing Pictorials	5
	M--Maintaining and Caring for Tools	20
	N--Reproducing Drawings	10
I	F--Detailing Single Parts	5
	G--Drawing Layouts	10
	J--Drawing Layouts of Single Parts	5
	L--Drawing Pictorials	5
	M--Maintaining and Caring for Tools	15
	N--Reproducing Drawings	10
J	F--Detailing Single Parts	25
	G--Drawing Layouts	20
	J--Drawing Layouts of Single Parts	25
	L--Drawing Pictorials	10
	M--Maintaining and Caring for Tools	15
	N--Reproducing Drawings	5
K	F--Detailing Single Parts	25
	G--Drawing Layouts	25
	H--Drawing Charts	5
	I--Drawing Diagrams	5
	J--Drawing Layouts of Single Parts	15
	L--Drawing Pictorials	5
	M--Maintaining and Caring for Tools	25
	N--Reproducing Drawings	15

Table 3.1.--Continued.

Pilot Site	Duty Code & Description	# of Items Requested
	Machine Trades	
A	D--Performing Metalwork Operations	10
	E--Performing Bench Work	10
	F--Operating Drill Presses	20
	G--Operating Grinding Machines	15
	H--Lathe Operations	25
	I--Operating Milling Machines	10
	J--Operating Power Saws	5
	O--Maintaining Shop Facilities and	5
	Work Areas	
B	A--Performing Supervisory Functions	5
	B--Performing Mathematical Calculations	5
	C--Designing and Planning Machine Work	15
	D--Performing Metalwork Operations	5
	E--Performing Bench Work	15
	F--Operating Drill Presses	10
	H--Lathe Operations	25
	I--Operating Milling Machines	15
	O--Maintaining Shop Facilities and	5
	Work Areas	
C	C--Designing and Planning Machine Work	5
	D--Performing Metalwork Operations	5
	E--Performing Bench Work	5
	F--Operating Drill Presses	15
	G--Operating Grinding Machines	15
	H--Lathe Operations	25
	I--Operating Milling Machines	25
	J--Operating Power Saws	
D	C--Designing and Planning Machine Work	5
	D--Performing Metalwork Operations	5
	E--Performing Bench Work	10
	F--Operating Drill Presses	15
	G--Operating Grinding Machines	15
	H--Lathe Operations	20
	I--Operating Milling Machines	25
	J--Operating Power Saws	5

Table 3.1.--Continued.

Pilot Site	Duty Code & Description	# of Items Requested
E	A--Performing Supervisory Functions	5
	B--Performing Mathematical Calculations	5
	C--Designing and Planning Machine Work	5
	D--Performing Metalwork Operations	5
	E--Performing Bench Work	5
	F--Operating Drill Presses	10
	G--Operating Grinding Machines	10
	H--Lathe Operations	20
	I--Operating Milling Machines	20
	J--Operating Power Saws	5
	L--Operating Shapers	5
	N--Maintaining Machines and Tools	5
	O--Maintaining Shop Facilities and Work Areas	5
F	B--Performing Mathematical Calculations	5
	C--Designing and Planning Machine Work	10
	D--Performing Metalwork Operations	10
	E--Performing Bench Work	10
	F--Operating Drill Presses	15
	G--Operating Grinding Machines	15
	H--Lathe Operations	30
	I--Operating Milling Machines	30
	J--Operating Power Saws	5
G	B--Performing Mathematical Calculations	10
	C--Designing and Planning Machine Work	5
	D--Performing Metalwork Operations	10
	E--Performing Bench Work	5
	F--Operating Drill Presses	10
	G--Operating Grinding Machines	5
	H--Lathe Operations	20
	I--Operating Milling Machines	20
	J--Operating Power Saws	
	O--Maintaining Shop Facilities and Work Areas	5

Table 3.2.--Number of items in each duty area available for inclusion in tests.

Duty Code	Description of Duty	# of Items Available
	Mechanical Drafting	
A	Designing Assembly Tooling	35
B	Designing Cutting Dies	3
C	Designing Gages	3
D	Designing Molds	11
E	Designing Specialty Fixtures	82
F	Detailing Single Parts	195
G	Drawing Layouts	184
H	Drawing Charts	49
I	Drawing Diagrams	71
J	Drawing Layouts of Single Parts	113
K	Drawing Machine Diagrams	27
L	Drawing Pictorials	59
M	Maintaining and Caring for Tools	255
N	Reproducing Drawings	48
O	Trouble-Shooting and Correcting Design Errors	6
	Total Written Items =	1,141
	Machine Trades	
A	Performing Supervisory Functions	12
B	Performing Mathematical Calculations	19
C	Designing and Planning Machine Work	28
D	Performing Metalwork Operations	59
E	Performing Bench Work	21
F	Operating Drill Presses	79
G	Operating Grinding Machines	80
H	Lathe Operations	235
I	Operating Milling Machines	122
J	Operating Power Saws	30
K	Operating Presses	3
L	Operating Shapers	28
M	Performing Production Machinist Line	3
N	Maintaining Machines and Tools	65
O	Maintaining Shop Facilities/Work Areas	9
	Total Written Items =	784

Chapter I, the rankings were artificial dichotomies (pass/fail), collected at the duty level. These estimates were correlated with the artificially dichotomized student subscores (pass/fail) obtained on the specific subtests addressed in the vocational/occupational assessment test. Student subscores were placed into artificial dichotomies because of the relatively few participants at each site.

The instructor ranking sheets included the maximum number of duty areas requested by the instructors. Therefore, not all V-TECS duties were used in constructing the ranking sheets. (For consensus agreements, see Table 3.1.) On the sheets, each duty area was preceded by a simple pass/fail box for the instructors to check, indicating their ranking of a particular student for that duty area. This system allowed instructors to indicate quickly whether or not the student possessed the skill/ knowledge in each of the duty areas tested. The pass-fail criterion was as follows. For a "pass" the student was judged to possess enough skill/knowledge to obtain a ranking of 50% or higher on questions pertaining to a specific skill (duty) area. Students receiving a "fail" were judged not to possess enough skill/ knowledge in the stated skill (duty) area to obtain a ranking of 50%.

The instructor ranking sheets required one additional item of information, the students' percentile ranking. To elicit meaningful rankings quickly, the researcher divided the 100 points of the percentile scale into five equal sectors (1-5) of 20 percentage points each (sector one = 0-20 points, sector two = 20-40 points, and so on). The instructors could easily rate students within one of the five sectors.

This criterion measure was then correlated with the percentage ranking obtained by students on the vocational/occupational assessment tests.

The instructor ranking sheets were pilot tested in two stages. First, each instructor reviewed the format of these sheets during inservice meetings conducted at various sites throughout Michigan during March and April 1986. Second, vocational curriculum specialists, as well as research and testing professionals, critiqued the ranking sheets' construction and format. The researcher considered all opinions before constructing the final forms for the study. (See Appendix E for the final version of the instructor ranking sheets.)

Data Collection

The instructor ranking sheets were mailed or delivered to the two groups of pilot-site instructors--mechanical drafting and machine trades. Included was a cover letter with directions for completing and returning the instruments, as well as a stamped, addressed envelope in which participants were to return the completed sheets. (See Appendix D.) Each pilot-site instructor received a ranking sheet for each student in the class. The following information was already on the top of the sheet to insure proper coding: instructor code, student number/name, program code/title, and school/center code. The researcher conducted a telephone follow-up shortly after the testing and retesting had taken place. All the instructor ranking sheets were returned.

During May, June, and July 1986, the researcher conducted a series of test-retests with the students from the 11 vocational/

occupational programs described earlier in this chapter. Each site included in the study was given a test and a retest date; thus 22 site visitations were required for data collection. The researcher conducted 20 of the 22 test/retest administrations of the assessment instrument. Twice he could not be present for either the test or the retest because of student motivation problems and/or time conflicts with other pilot-test locations. In both cases, the pilot-site instructors were well informed about the testing/retesting process and the proper administration of the instruments.

The researcher used the following procedures with the students in both the test and retest phases:

1. Introduced the test-administration process to the students.

2. Discussed with students the reasons for the pilot testing/retesting effort.

3. Gave students instructions for filling in the computer-scan answer sheet.

4. Encouraged the students to do their best and wished them well on the examination.

On the test day, Form A2 of the vocational/occupational student assessment test was administered to all students in attendance. Students were given enough time to complete all items on the test. They were encouraged to take their time and were told that the researcher would stay after the scheduled class session if they needed extra time to complete the test. However, this was not necessary because all students finished their tests within the scheduled time.

The retest of the vocational/occupational student assessment test was conducted at all 11 test sites within five to seven days of the first administration. Some students who had taken part on test day did not show up for retest day; conversely, some students not involved on test day were present on retest day. Students were allowed to take the test the one day they attended class; however, only data from students in attendance on both days (test and retest) were used to answer the research questions.

On the retest day, both forms of the vocational/occupational education assessment test were administered. As students entered the room or after they were seated, the test administrator handed out the test booklets and answer sheets. The tests were handed out from the top of a preordered pile of exams, which alternated between Form A2 and Form B2.

The students who took Form A2 of the assessment instrument on both occasions provided (in total or in part) the data used to answer Research Questions 1, 3, 5, and 6. The students who took Form A2 on test day and Form B2 on retest day provided (in total or in part) the data required to answer Research Questions 2, 4, 5, and 6.

Data Processing and Analysis

Once the testing period was over and both test and retest student answer sheets had been collected, the scoring process began. Header sheets and answer keys for the 22 forms (A and B) of the assessment instrument were prepared and taken to the Michigan State University computer scoring laboratory for processing. Full-item analysis statistics were requested for each student, by school and by program.

Data gathered from the instructor ranking sheets were entered by hand into a computer base and cross-checked by personnel in the Michigan State University Department of Applications Programming. The pass/fail rankings and percentile rankings were catalogued as the test-retest student data (by test site, program, and student number). Tables were constructed using the aforementioned data and delivered to the Department of Applications Programming of the Michigan State University computer laboratory for analysis. (See Figure 3.2.)

To answer the six research questions, the following methods were used. The students who took Form A2 of the assessment test on both occasions provided the data used to answer Research Question 1. Since the test data were in the form of continuous scores, the Pearson product-moment correlation (r) was computed to obtain a reliability estimate of test stability.

The students who took Form A2 of the assessment test on test day and Form B2 of the instrument on retest day provided the data used to answer Research Question 2. Again, since the test data were in the form of continuous scores, the Pearson product-moment correlation (r) was used to gain the reliability estimates of test stability and equivalence.

Test A2 and retest A2 scores were subdivided by duty area, which yielded a subscore for each duty tested on both occasions. These subtest scores were converted into artificial dichotomies (pass/fail)

Data table key

Student Number	Score Test A	Score Retest A	Score Retest B	Difference Between A/B	Percentage Rank Actual	Precentage Rank Predicted	Letter Grade	Average Duty Ranking Test-Retest Pass/Fail	
								Duty Rank on Test Pass/Fail	Duty Rank on Retest Pass/Fail
								Instructor Predicted Duty Ranking Pass/Fail	

Site A

Student	Score Test A	Score Retest A	Score Retest B	Difference	% Rank Actual	% Rank Predicted	Letter Grade	D	E	F	G	H	I	J	S
								P	P	F	P	P	P	P	P
001	51		58	7	5	5	D	P P	P P	F F	F P	P P	P P	P P	P P
								P	P	F	F	F	F	P	P
								F	F	F	F	F	F	P	F
002	37		45	8	2	4	C+	P F	F F	F F	F P	F P	F F	P P	F P
								P	P	F	P	P	F	P	P
								P	F	F	F	F	F	F	F
003	37		34	-3	2	4	C-	F P	P F	F F	F P	F F	F F	P F	F P
								P	P	F	F	P	P	P	P
								P	F	F	F	F	F	P	F
004	47	37		-10	3	4	C	P P	F F	F F	F P	F F	F F	F P	F P
								P	P	F	F	F	F	P	P
								F	P	F	F	P	F	F	P
005	28		40	12	2	4	C	F P	P F	F F	P F	F P	F F	F F	P P
								P	F	F	F	F	F	P	P
								F	F	F	F	F	F	F	P
006	27		34	7	1	3	C+	F F	F P	F F	F F	F F	F F	F F	P P
								F	F	F	F	F	F	P	P
								P	P	F	F	F	F	P	F
007	37	44		7	3	4	C+	P P	P F	F F	F P	F F	F F	P P	F F
								P	P	F	P	F	P	P	P
								F	F	F	F	F	F	P	P
008	40	47		7	4	4	C+	P F	P F	F F	P F	F F	F F	F P	P P
								P	P	F	P	P	P	P	P

Figure 3.2: Sample data table.

and correlated with each other using the tetrachoric correlation (r_t) to give a reliability estimate of stability within duty areas. When empty cells occurred in the 2 x 2 correlational matrix, the formula could not be used. Therefore, percentage agreements were computed between test and retest subscore rankings to help clarify the findings. These data provided the answer to Research Question 3.

Test A2 and retest B2 scores were also subdivided by duty and converted to artificial dichotomies, which yielded another set of dichotomous data to be correlated using the tetrachoric correlation (r_t) for reliability estimates of stability and equivalence. Again, when empty cells occurred in the 2 x 2 correlational matrix, the formula could not be used. Therefore, percentage agreements were computed between test and retest subscore rankings to help clarify the findings. These data were used to answer Research Question 4.

Student test scores were converted to percentile ranks. These rankings enabled the researcher to identify each subject's performance relative to the group of participants from his/her test site. The test site instructors also provided individual student rankings, on the instructor ranking sheets. The relationship that existed between these two sets of data was measured using the chi-square test. The results were computed for each site, for each occupational division, and across all sites. These results provided the answer to Research Question 5.

To answer Research Question 6, the researcher averaged each student's individual duty subscores from both the test and retest. This resulted in one subscore for each duty area addressed on the tests. The relatively small number of participants necessitated converting these subscores into artificial dichotomies (pass/fail). The test site instructors provided individual duty rankings for each participant. The relationship that existed between these two sets of data was computed using tetrachoric correlations for each duty area. The results were computed for each site, for each occupational division, and across all sites. The data correlated here were used to answer Research Question 6.

For all analyses, an alpha level of .05 or less was established to indicate statistical significance. Results of the data analyses are presented in Chapter IV.

CHAPTER IV

RESULTS OF THE DATA ANALYSIS

Introduction

Results of the analysis of data collected to answer the six research questions are presented in this chapter. Restated, the questions are as follows:

> Research Question 1: How stable are test scores obtained on the vocational/occupational assessment test?
>
> Research Question 2: How stable and equivalent are scores obtained on parallel forms of the vocational/occupational assessment test?
>
> Research Question 3: How stable are subscore rankings obtained on the vocational/occupational assessment test?
>
> Research Question 4: How stable and equivalent are subscore rankings obtained on parallel forms of the vocational/occupational assessment test?
>
> Research Question 5: How valid are the percentile rankings obtained on the vocational/occupational assessment test?
>
> Research Question 6: How valid are the subscore rankings obtained on the vocational/occupational assessment test?

The data sources for this study were (a) the test/retest scores of selected Michigan vocational/occupational students involved in the occupational assessment pilot-testing program conducted during spring 1986 and (b) instructors' rankings of students' knowledge, also collected in spring 1986. The occupational specialties included in this research were mechanical drafting and machine trades. Seven test sites were used in machine trades and four in mechanical drafting (machine trades test site G was eliminated from the analysis due to low student participation).

In the following pages, the findings for both occupational specialties are presented according to the research questions posed. Findings are presented by occupational specialty relative to the research questions.

Results

Research Question 1

> How stable are test scores obtained on the vocational/occupational assessment test?

Machine trades. The Pearson product-moment coefficient of correlation (r) for the test/retest scores on Form A2 of the machine trades assessment test was computed for each site (see Table 4.1). The r's for three of the six sites (Sites C, E, and F) were significant at the .05 level. According to Rowntree (1981), the relationships found for Sits A, C, E, and F were either moderate or strong. Even though Rowntree's definition denotes moderate to strong relationships, the reader is advised to interpret the findings cautiously because of the small population size.

Mechanical drafting. Table 4.2 contains the results for the analysis of the mechanical drafting test/retest scores. The Pearson r's for the test/retest scores were significant at the .05 level for three of the four test sites (Sites I, J, and K). The relationships were moderate to very strong (Rowntree, 1981). Again, given the small population size, the reader is cautioned about the practical significance of the results.

Table 4.1.--Correlational data for machine trades test Form A2/retest Form A2.

Site	Pearson r Value
A	.590 (n = 5)
B	.116 (n = 16)
C	.820* (n = 13)
D	.371 (n = 10)
E	.687* (n = 7)
F	.723* (n = 9)

*Significant at the .05 level.

Table 4.2.--Correlational data for mechanical drafting test/retest, Form A2.

Site	Pearson r Value
H	.034 (n = 18)
I	.548* (n = 11)
J	.917* (n = 7)
K	.763* (n = 7)

*Significant at the .05 level.

Overall, the data pertaining to the measures of stability were moderate. Significant relationships were found at six of the ten sites (Sites C, E, F, I, J, and K), and at one other site (Site A) a moderate relationship was found. A Pearson r of .917 for mechanical drafting Site J was the highest, whereas the r of .034 computed for Site H was the lowest. The results of the analysis seem to indicate stability in the machine trades and mechanical drafting test scores.

Research Question 2

> How stable and equivalent are scores obtained on parallel forms of the vocational/occupational assessment test?

Machine trades. Table 4.3 contains the Pearson r's for the test/retest scores on parallel forms of the machine trades assessment test. Strong relationships were found between sets of student scores on test Forms A2 and B2 at five of the six sites (Sites A, C, D, E, and F). The results at Sites C, D, E, and F were significant at the .01 level, whereas those at Site A were significant at the .05 level. These results suggest that Forms A2 and B2 of the machine trades assessment test are equivalent and that the scores appear to be stable.

Table 4.3.--Correlational data, machine trades test Form A2/retest form B2.

Site	Pearson r Value
A	.817* (n = 5)
B	.280 (n = 21)
C	.710** (n = 12)
D	.852** (n = 11)
E	.812** (n = 8)
F	.724** (n = 13)

*Significant at the .05 level.

**Significant at the .01 level.

Mechanical drafting. As indicated in Table 4.4, positive Pearson r's were found between test/retest scores on Forms A2 and B2 of the mechanical drafting assessment test. Three of the four test sites (Sites I, J, and K) had results that were significant at the .05 level. Even though the relationships can only be described as moderate (Rowntree, 1981), the mechanical drafting assessment tests seem to be equivalent, and scores on those tests appear to be stable.

Table 4.4.--Correlational data for mechanical drafting test Form A2/ retest Form B2.

Site	Pearson r Value
H	.071 (n = 21)
I	.592* (n = 10)
J	.566* (n = 11)
K	.661* (n = 12)

*Significant at the .05 level.

The relationship between the scores on equivalent forms of the vocational/occupational assessment tests was significant at eight of the ten sites. Even though the relationships were moderate, the results were significant and appear to indicate the equivalence and stability of the tests and scores, respectively. That is, the tests can be used for both long-range prediction and inferences to a domain of knowledge.

Research Question 3

How stable are subscore rankings obtained on the vocational/occupational assessment test?

Machine trades. Results of the test/retest of machine trades assessment test Forms A2, across all sites, are shown in Table 4.5. The tetrachoric correlations computed for each site indicated moderate to very strong relationships between students' subscore rankings on the test and their subscore rankings on the retest. The highest tetrachoric correlation (.737) was found at Site C; the lowest tetrachoric correlation was found at Site D (.124). At four of the machine trades sites (Sites A, C, E, and F) where students were tested and retested with the same form (A2) of the assessment instrument, moderate relationships (Rowntree, 1981) were noted between subscore rankings. Even though the relationships were moderate, at all four sites they were significant at the .05 level or lower. The results imply there was stability between test/retest student subscore rankings.

Mechanical drafting. Table 4.6 contains the results of the analysis for mechanical drafting. The combined mean percentage agreements indicated relationships ranging from moderate to strong. Site H had the largest percentage agreement (94%), although a tetrachoric correlation statistic could not be computed. At three of the four sites (Sites I, J, and K) where Form A2 of the mechanical drafting test was administered twice, the relationships were significant at the .01 level. These results seem to indicate stability within students' subscore rankings on the mechanical drafting assessment tests.

Table 4.5.--Subscore ranking agreement and tetrachoric correlations: machine trades, test/retest Form A2.

Site	Mean Percentage Agreement	Mean Percentage Disagreement	Tetrachoric Correlation (r_t)
A	73	27	.583* (n = 40)
B	70	30	.134 (n = 96)
C	77	23	.737** (n = 104)
D	54	46	.124 (n = 70)
E	67	33	.47/* (n = 56)
F	78	22	.733** (n = 54)

Note: n = students x subtests (duty areas).

*Significant at the .05 level.

**Significant at the .01 level.

Table 4.6.--Subscore ranking agreement and tetrachoric correlations: mechanical drafting, test/retest Form A2.

Site	Mean Percentage Agreement	Mean Percentage Disagreement	Tetrachoric Correlation (r_t)
H	94	6	---
I	71	29	.548* (n = 60)
J	74	26	.640* (n = 66)
K	88	12	.850* (n = 96)

Note: n = Students x subtests (duty areas).

*Significant at the .01 level.

The tetrachoric correlations for subscore stability are shown in Table 4.7 for both occupational specialties. The tetrachoric correlations were both significant at the .001 level. Mechanical drafting had the largest reported statistic (.871), whereas machine trades had the smallest (.561). The results indicated that students' subscores on the machine trades and mechanical drafting assessment tests were stable over time.

Table 4.7.--Tetrachoric correlations for combined subscore rankings on the vocational/occupational assessment test.

	All Subscore Rankings Test A2/Retest A2
Machine trades	.561* (n = 330)
Mechanical drafting	.871* (n = 222)

Note: n = students x subtests (duty areas).

*Significant at the .001 level.

Research Question 4

How stable and equivalent are subscore rankings obtained on parallel forms of the vocational/occupational assessment test?

Machine trades. The degree of relationship found between duty subscore rankings on parallel forms of the machine trades assessment test is shown in Table 4.8. All subscore rankings were combined by site and averaged to compute a tetrachoric correlation for each site. At two of the sites (Sites C and E), the relationships were significant. At the other sites, nonsignificant relationships were found. Site E had the highest percentage agreement (71%), whereas Site D had the lowest agreement between test/retest subscore rankings (54%).

Table 4.8.--Subscore ranking agreement and tetrachoric correlations: machine trades, test Form A2/retest Form B2.

Site	Mean Percentage Agreement	Mean Percentage Disagreement	Tetrachoric Correlation (r_t)	
A	63	37	.387	(n = 40)
B	64	36	.027	(n = 96)
C	60	40	.280*	(n = 104)
D	54	46	.121	(n = 70)
E	71	29	.596**	(n = 56)
F	61	39	.196	(n = 54)

Note: n = students x subtests (duty areas).

*Significant at the .05 level.

**Significant at the .01 level.

Mechanical drafting. The percentage of subscore ranking agreement on equivalent forms of the mechanical drafting test is shown in Table 4.9. The relationships recorded at Sites I and J were weak; however, the tetrachoric correlation at Site J was significant at the .05 level. Site H had the highest percentage agreement for the mechanical drafting sites (88%). Site K had the lowest percentage agreement (58%), followed by Site I (55%).

The results of the tetrachoric correlations for both occupational specialties with regard to subscore stability and equivalence are shown in Table 4.10. The tetrachoric correlations were both significant at the .001 level. Mechanical drafting had the strongest reported relationship (.532), whereas machine trades had the weakest (.333). The results of the analysis suggest that students' subscores on the machine trades and mechanical drafting assessment tests were stable and that the subtests (duty areas) were equivalent.

Table 4.9.--Subscore ranking agreement and tetrachoric correlations: mechanical drafting, test Form A2/retest Form B2.

Site	Mean Percentage Agreement	Mean Percentage Disagreement	Tetrachoric Correlation (r_t)
H	88	12	---
I	55	45	.351 (n = 60)
J	73	27	.336* (n = 66)
K	58	42	.136 (n = 96)

Note: n = students x subtests (duty areas).

*Significant at the .05 level.

Table 4.10.--Tetrachoric correlations for combined subscore rankings on the vocational/occupational assessment test.

	All Subscore Rankings Test A2/Retest B2
Machine trades	.333* (n = 330)
Mechanical drafting	.532* (n = 222)

Note: n = students x subtests (duty areas).

*Significant at the .001 level.

Research Question 5

How valid are the percentile rankings obtained on the vocational/ occupational assessment test?

Machine trades. The percentage of agreement between students' achievement on the tests and instructors' rankings of the students across all sites is displayed in Table 4.11. In 30% of the cases, student performance matched instructors' rankings. More than half (52.5%) of the machine trades students did not achieve as well as their instructors expected they would. Seventeen percent of the students performed better than expected.

Table 4.11.--Students' percentile rank achievement in relation to instructors' rankings of students: machine trades, all sites combined, in frequency and percentage of total (N = 130).

Actual Rank	Instructor Ranking 1		2		3		4		5	
	No.	% of Total	No.	% of Total	No.	% of Total	No.	% of Total	No.	% of Total
1	1	*.5	7	-5.0	7	-5.0			2	-1.0
2	3	+2.0	6	*4.5	11	-8.0	9	-7.0	5	-4.0
3			2	+1.0	8	*6.0	13	-10.0	5	-4.0
4			1	+ .5	7	+5.0	12	*9.0	9	-7.0
5					4	+3.0	8	+6.0	13	*10.0

Key: *Percentage agreement = 30%.
+Student achievement above ranking = 17.5%.
-Student achievement below ranking = 52.5%.

As shown in Table 4.12, the percentage agreement between machine trades students' percentile rankings based on test achievement and their instructors' rankings was weak. However, the chi-square value was significant at four sites (Sites A, C, D, and F). The highest agreement between rankings was recorded at Site E (53.4%); the lowest was recorded at Site B (18.9% match between students' achievement and instructors' rankings). On the average, 48% of the machine trades students performed at a rate below what their instructors expected. Conversely, 19.5% of the machine trades students achieved better than their instructors expected. These results seem to indicate little relationship between students' percentile rank achievement and their instructors' rankings of them.

Table 4.12.--Percentages and chi-square results for students' rank achievement in relation to instructors' rankings of students' achievement: machine trades.

Site	Instructor Ranking/Student Rank			Chi-Square Results	
	% Agreement	% Above	% Below		
A	30.0	0	70.0	20.0**	(n = 10)
B	18.9	29.7	51.4	18.9	(n = 37)
C	44.0	24.0	32.0	26.1*	(n = 25)
D	19.0	0	81.0	19.4*	(n = 21)
E	53.4	20.0	26.6	21.3	(n = 15)
F	27.2	9.1	63.7	37.5**	(n = 22)

*Significant at the .05 level.

**Significant at the .01 level.

Mechanical drafting. The overall percentage of agreement between students' achievement on the tests and instructors' rankings of their students is illustrated in Table 4.13. In 30.9% of the cases, the mechanical drafting students' actual quintile rank matched the ranking provided by the site instructors. Forty-four percent of the students fell short of their instructors' expectations, and 25% achieved better than expected.

Table 4.13.--Students' percentile rank achievement in relation to instructors' rankings of students: mechanical drafting, all sites combined, in frequency and percentage of total (N = 97).

	Instructor Ranking									
Actual Rank	1		2		3		4		5	
	No.	% of Total	No.	% of Total	No.	% of Total	No.	% of Total	No.	% of Total
1	4	*4	2	-2	2	-2	3	-3	1	-1
2	3	+3	3	*3	5	-5	10	-10	3	-3
3	2	+2	7	+7	6	*6	7	-7	5	-5
4	2	+2	1	+1	5	+5	10	*10	4	-4
5					2	+2	3	+3	7	*7

Key: *Percentage agreement = 30.9%.
+Student achievement above ranking = 25%.
-Student achievement below ranking = 44.1%.

Illustrated in Table 4.14 is the percentage agreement between students' rankings on the mechanical drafting assessment test and instructors' rankings of their students. Results of the chi-square analysis are also presented. Overall, the results appear to indicate a poor relationship between the two variables tested.

Table 4.14.--Percentages and chi-square results for students' rank achievement in relation to instructors' rankings of students' achievement: mechanical drafting.

Site	Instructor Ranking/Student Rank			Chi-Square Results	
	% Agreement	% Above	% Below		
H	33.3	53.8	12.9	19.0	(n = 39)
I	14.3	9.6	76.1	5.9	(n = 21)
J	50.1	0	49.9	27.2*	(n = 18)
K	26.3	10.6	63.1	4.3	(n = 19)

*Significant at the .01 level.

Research Question 6

How valid are the subscore rankings obtained on the vocational/occupational assessment test?

Machine trades. Shown in Table 4.14 are the frequencies and percentages of the instructors' perceived rankings and students' actual subscore rankings in four cells: pass/pass, pass/fail, fail/pass, and fail/fail. Fifty-eight percent of the time (33% + 25%), instructors predicted accurately (by ranking each student on each duty area covered in the test) students' achievement (pass/pass or fail/fail). Forty-two percent of the time, the students' performance was above or below their instructors' rankings. Ten percent of the students fared better than expected, whereas 32% achieved at a rate lower than their instructors' rankings.

Table 4.15.--Vocational/occupational assessment test subscore averages in relation to instructors' rankings: machine trades, all sites.

Students' Subscore Rankings		Instructors' Rankings	
		Pass	Fail
Pass	No.	308	84
	%	+33	-10
Fail	No.	302	231
	%	-32	+25

Note: +58% agreement.
-42% disagreement.
The relationship between percentage agreements was significant at the .01 level. N = 925 (students x subscore rankings).

Table 4.16 contains the results of the data analysis for machine trades, by test site and for all sites combined. The tetrachoric correlation between instructors' rankings and students' subscore rankings was statistically significant at all sites except Site B. The tetrachoric correlations for Sites D, E, and F were significant at the .001 level, whereas the correlations for Sites A and C were significant at the .05 level. The tetrachoric correlation for all machine trades sites combined was significant at the .001 level. The analysis results appeared to indicate that relationships between students' achieved percentile rankings and instructors' rankings of their students' achievement were valid. However, because of the small population size, the reader is cautioned about the practical significance of the results.

Table 4.16.--Instructors' rankings in relation to students' subscore rankings: percentage agreement and tetrachoric correlation data, machine trades.

	Instructors' Rankings/Subscore Rankings			
Site	Mean Percentage Agreement	Mean Percentage Disagreement	Tetrachoric Correlation (r_t)	
A	59	41	.332*	(n = 10)
B	45	65	.005	(n = 37)
C	55	45	.227*	(n = 25)
D	64	36	.532**	(n = 21)
E	79	21	.490**	(n = 15)
F	77	23	.679**	(n = 22)
A-F	58	42	.371**	(n = 130)

*Significant at the .05 level.

**Significant at the .01 level.

Mechanical drafting. Illustrated in Table 4.17 are the frequencies and percentages of instructors' subscore rankings of students and students' actual subscore rankings in four cells: pass/pass, pass/fail, fail/pass, and fail/fail. Sixty-nine percent of the time, students who achieved "pass" rankings on duty areas were ranked by their instructors as possessing the knowledge to perform at a passing level. Twenty percent of the time, students who achieved "pass" rankings on duty areas were rated by their instructors as possessing the knowledge to perform at a passing level. Thirty-one percent of the time, instructors' rankings and students' subscore rankings did not match; in 19% of these instances, instructors ranked students as passing but the students actually failed the duty area.

Table 4.17.--Vocational/occupational assessment test subscore averages in relation to instructors' rankings: mechanical drafting, all sites.

Students' Subscore Rankings		Instructors' Rankings Pass	Instructors' Rankings Fail
Pass	No.	121	75
	%	+33	-10
Fail	No.	120	304
	%	-19	+49

Note: +69% agreement.
-31% disagreement.
The relationship of percentage agreements was significant at the .001 level. N = 620 (students x subscore rankings).

Table 4.18 contains the results of correlational analyses performed on the mechanical drafting data, by site and for all sites combined. The tetrachoric correlations revealed a weak relationship; only one test site (Site I) had a correlation that was significant at the .05 level. The correlational result for all mechanical drafting sites combined was significant at the .001 level. The individual-site data illustrated a poor relationship between instructors' rankings of student achievement (per subscore rank) and actual student subscore achievement. However, combining all mechanical drafting site data suggested a statistically significant relationship between the test and the criterion measure. Again, the practical significance of this finding must be viewed with caution.

Table 4.18.--Instructors' rankings in relation to students' subscore rankings: percentage agreement and tetrachoric correlation data. mechanical drafting.

	Instructors' Rankings/Subscore Rankings			
Site	Mean Percentage Agreement	Mean Percentage Disagreement	Tetrachoric Correlation (r_t)	
H	99	1	---	
I	53	47	.259*	(n = 21)
J	54	46	.185	(n = 17)
K	86	14	.072	(n = 19)
H-K	69	31	.494**	(n = 57)

*Significant at the .05 level.

**Significant at the .001 level.

Summary

The results of the data analysis provided answers to the research questions regarding the reliability and validity of the vocational/occupational assessment tests. Strong to moderate estimates of stability were reported for the vocational/occupational assessment tests. Stability and equivalence estimates for the tests were also moderate to strong; statistically significant results were found at 80% of the test sites. Subscore rankings yielded statistically significant results 78% of the time, suggesting stability of subscores. Stability and equivalence estimates were significant 30% of the time within individual sites; these estimates were significant at the .001 level for all sites combined. Tetrachoric correlations and chi-square values computed between the tests and the criterion measures were statistically significant at 50% of the test sites. The mechanical drafting group had the weakest relationships between students' achievement (percentage ranking on the test) and their instructors' rankings (criterion measure). The test's concurrent validity was established by correlating students' subscore rankings with instructors' rankings of their students using chi-square and tetrachoric correlation. The result was statistically significant at the .001 level.

A summary of the findings is presented in Chapter V, along with conclusions drawn from those findings and recommendations for further research.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

Summary

The researcher's purpose in this study was to obtain estimates of reliability and measures of validity on assessment tests developed using the occupational test-item bank at Michigan State University. The occupational specialties selected for this study were machine trades and mechanical drafting. The following research questions were investigated:

> Research Question 1: How stable are test scores obtained on the vocational/occupational assessment test?

> Research Question 2: How stable and equivalent are scores obtained on parallel forms of the vocational/occupational assessment test?

> Research Question 3: How stable are subscore rankings obtained on the vocational/occupational assessment test?

> Research Question 4: How stable and equivalent are subscore rankings obtained on parallel forms of the vocational/occupational assessment test?

> Research Question 5: How valid are the percentile rankings obtained on the vocational/occupational assessment test?

> Research Question 6: How valid are the subscore rankings obtained on the vocational/occupational assessment test?

<u>Procedures</u>

The population was composed of 171 machine trades students and 117 mechanical drafting students (N = 288) from 11 vocational/occupational education programs representing the machine trades and mechanical drafting occupational specialties, and the program instructors who volunteered to participate in the study. The data analyzed for the reliability portion of the study were the usable test/retest scores of 230 vocational/occupational students.

For each occupational specialty, subjects were initially tested with Form A2 of the vocational/occupational assessment test (developed by their instructors in conjunction with Michigan State University test-item-bank personnel). From five to seven days after taking the test, the students were retested with either Form A2 or a parallel form (B2) of the vocational/occupational assessment test. Stability and stability and equivalence estimates were computed.

The test site instructors were asked to rank their students' knowledge at both the occupational and duty levels, using the Vocational-Technical Education Consortium of States (V-TECS) duty and task listings modified for this study. Data provided by the instructors constituted the criterion measure used in studying the concurrent validity of the assessment tests. Although content validity, per se, was not a part of this study, the manner in which the tests were constructed was taken as evidence that they were content valid.

Findings and Conclusions

The findings and conclusions are presented separately for the two occupational specialties. Rowntree's (1981) labels (strong, weak, and so on) were used to interpret the strengths of the relationships.

Machine Trades

1. Pearson product-moment coefficients of stability averaged 0.55 for the machine trades assessment tests, which was interpreted as moderate.

As sample size decreases, reliability estimates typically decrease as well (Ebel, 1979; Mehrens & Lehmann, 1984), making it increasingly difficult to obtain statistical significance. It is conceivable, because of the small number of participants, that the average stability coefficients for the machine trades tests were spuriously low. A larger sample might have resulted in a more heterogeneous group, thereby yielding a higher reliability coefficient. Although the stability coefficient obtained in this study was interpreted as moderate, the researcher is confident that one can generalize from the score a student receives at one time to what he/she would obtain had the test been administered at another time.

2. Pearson product-moment coefficients of stability and equivalence obtained on parallel forms of the machine trades assessment test averaged 0.70, which was interpreted as strong.

Typically, vocational/occupational teachers do not prepare alternate forms of achievement tests. However, Michigan State University's vocational/occupational test-item bank was used to produce parallel forms of assessment tests from occupationally specific subbanks. Seemingly, tests developed with items selected at random from the same subbank (of similar items) should be equivalent and should yield stable scores. This was the case with the machine trades assessment tests. Estimates of stability and equivalence are usually lower than stability estimates alone (Mehrens & Lehmann, 1984). This was not the case with parallel forms of the machine trades test. Based on this finding, the researcher concluded that examinees were as challenged in responding to Form B2 of the machine trades assessment test as they were in responding to similar items on Form A2. Therefore, the estimates of stability and equivalence are evidence that, given a different but similar set of items asked at a different time, students would obtain similar test results.

3. The tetrachoric correlation coefficient of stability obtained on students' subscore rankings (pass/fail) on the machine trades assessment test was 0.56, which was interpreted as moderate.

Again, as occurred when stability coefficients on the machine trades assessment test were computed, similar results were obtained when the students' subscore (duty area) rankings were used. The small population, as well as the relatively small number of items representing each subtest (duty area), may have caused the stability coefficients to be spuriously low. With a larger group, the tetrachoric correlation might have been larger, resulting in a "stronger" interpretation. Seventy percent of the machine trades students' subscore rankings remained consistent from one occasion to the next (test/retest). The tetrachoric correlation was significant at the .001 level. Based on this finding, the researcher is confident that one can generalize from the ranking (pass/fail) a student obtained on one occasion to what ranking he/she would obtain had the test been administered at another time.

4. The tetrachoric correlation coefficient of stability and equivalence obtained on students' subscore rankings (pass/fail) on parallel forms of the machine trades assessment test was 0.33, which was interpreted as weak.

Two factors cited in the literature as influencing test/retest reliability coefficients were (a) student answers on the retest are not totally independent of their answers on the initial test, and (b) general lack of interest on the students' part may make the retest a poor measure of his/her knowledge (Ebel, 1979). The tetrachoric correlation statistic used to obtain estimates of reliability (stability and equivalence) on subscore rankings obtained using parallel forms of the machine trades assessment test may well have been affected by both factors. Therefore, the reader is advised to interpret this finding regarding stability and equivalence of the machine trades subscore rankings with caution. Overall, the results were statistically significant at the .001 level, and it appears that, given a different but similar set of items asked at a different time, student subscore rankings would remain stable.

5. The chi-square value of 23.8 between students' percentile rankings obtained on the machine trades assessment test and the instructors' percentile rankings of students was significant at the .001 level and interpreted as strong.

This result, however encouraging, must be interpreted with caution. More than one-half of the students (51%) obtained a rank below what their instructors had given them; however, only 21% of those disagreed by more than 20 percentage points (see Appendix F, Table F.43). Based on this finding, the researcher concluded that there was little relationship between the test results (student percentile rank) and the criterion measure (instructors' rankings). In addition, this relationship between test and criterion measure was taken as evidence of the test's lack of concurrent validity.

6. The tetrachoric correlation between students' average subscore rankings (pass/fail) on the machine trades assessment test (test/retest) and the instructors' subscore rankings was 0.37, which was interpreted as weak.

One of the most common misinterpretations made in educational research is to confuse statistical significance with practical significance (Borg & Gall, 1983). Although statistically significant results were obtained at five of the six machine trades test sites (A, C, D, E, and F), there seemed to be little relationship, at the duty level, between instructors' rankings of students' knowledge (criterion measure) and students' actual rankings on the machine trades assessment test. The rankings achieved on the test matched the rankings provided by the instructors only 30% of the time. It is conceivable that the machine trades instructors selected students' rankings based on their observation/knowledge of the students' ability to perform machining tasks (psychomotor skill development), rather than on the students' knowledge of machine trades practices and procedures (cognitive development). The evidence indicated that the tests did not have concurrent validity.

Mechanical Drafting

1. Pearson product-moment coefficients of stability averaged 0.57 for the mechanical drafting assessment test, which was interpreted as moderate.

Three of the four mechanical drafting test sites (I, J, and K) had results that were statistically significant at the .05 level. It is conceivable that the low reliability estimate at Site H could be explained, in part at least, by the lack of variance in the students' test scores (70% of the students' scores ranged between 20 and 29 points). Eliminating Site H from the mechanical drafting group resulted in a Pearson product-moment stability coefficient of 0.74, which was interpreted as strong. Based on this finding, the researcher concluded that one can generalize from the score a student receives at one time to what he/she would obtain had the test been given at another time.

2. Pearson product-moment coefficients of stability and equivalence obtained on parallel forms of the mechanical drafting assessment test averaged 0.47, which was interpreted as moderate.

Although statistically significant results were obtained at three of the four mechanical drafting test sites (I, J, and K), the stability and equivalence estimates were considered moderate because of a very poor correlation at Site H ($r = 0.07$). If Site H had been eliminated, the correlations would have been "stronger." Also, estimates of stability and equivalence are usually lower than either stability or equivalence estimates alone (Mehrens & Lehmann, 1984). This was the case with parallel forms of the mechanical drafting assessment test. Therefore, the estimates of stability and equivalence indicated that, given a different but similar group of items administered at a different time, students would obtain similar results.

3. The tetrachoric correlation coefficient of stability obtained on students' subscore rankings (pass/fail) on the mechanical drafting assessment test was 0.87, which was interpreted as strong.

As occurred when stability coefficients on the mechanical drafting assessment tests were computed, similar but "stronger" results were obtained when the students' subscore (duty area) rankings were used. Eighty-one percent of the mechanical drafting students' subscore rankings remained consistent from one occasion to the next (test/retest). The tetrachoric correlation was significant at the .001 level. The researcher concluded that one can generalize from the ranking (pass/fail) a student received on one occasion to what ranking he/she would receive had the test been administered on another occasion.

4. The tetrachoric correlation coefficient of stability and equivalence obtained on students' subscore rankings (pass/fail) on parallel forms of the mechanical drafting assessment test was 0.53, which was interpreted as moderate.

The students' subscore rankings on parallel forms of the mechanical drafting assessment test (test Form A2/retest Form B2) were consistent 69% of the time. Overall, the tetrachoric correlation was statistically significant at the .001 level. Thus, the researcher is confident that, had a similar yet different set of items been asked at a different time, students' subscore rankings would have remained stable.

5. The chi-square value of 14.1 between students' percentile rankings obtained on the mechanical drafting assessment test and the instructors' percentile rankings of students was not significant and was interpreted as weak.

Overall, 69.1% of the rankings obtained by students on the mechanical drafting test differed from the rankings provided by their instructors. These results seem to indicate that instructors who teach and assess students primarily in the psychomotor domain may be ill-equipped (poor criterion measure) to rank or assess students in the cognitive and affective domains. Based on this finding, the researcher concluded that there was little relationship between the test results (ranking of students by percentage rank) and the criterion measure (instructors' rankings of students). The small relationship (between test and criterion measure) was taken as evidence that the test lacked concurrent validity.

6. The tetrachoric correlation between students' average subscore rankings (pass/fail) on the mechanical drafting assessment test and the instructors' subscore rankings was 0.49, which was interpreted as moderate.

Again, a common misinterpretation is to confuse statistical significance with practical significance (Borg & Gall, 1983). Although statistically significant tetrachoric correlation coefficients were computed using the combined mechanical drafting data, only one mechanical drafting test site (Site I) had a significant tetrachoric correlation between students' actual subscore rankings and their instructors' subscore rankings of them. It is conceivable, as with the machine trades group, that instructors selected test items based on their observation or knowledge of students' ability to perform drafting skills, rather than their students' knowledge of drafting practices and procedures. Including more test items in each subtest (duty area) might have increased the test's content validity. However, the results indicated moderate agreement between the students' subscore rankings (the test) and the instructors' subscore rankings of students (the criterion measure). Based on the evidence, the researcher concluded that the test had little concurrent validity.

Synthesis

Although findings and conclusions were presented separately for each occupational specialty, the reader should note the similarities between the reliability and validity estimates of the assessment instruments developed for machine trades students and those developed for mechanical drafting students. The coefficients of stability and of stability and equivalence obtained on the occupational assessment instruments were statistically significant for both groups. Additionally, coefficients of stability and of stability and equivalence using the subscore rankings were also statistically significant for each group.

The chi-square test indicated a statistically significant relationship between the machine trades students' percentile ranking obtained on the tests and their instructors' rankings of them. However, the chi-square test using data collected from the mechanical drafting students and their instructors failed to produce a statistically significant result. It is possible that the mechanical drafting instructors ranked students according to their mechanical drafting skills rather than their cognitive understanding of drafting practices and procedures.

The tetrachoric correlations between the students' subscore rankings and their instructors' rankings indicated that the tests had concurrent validity. Although the results of the tetrachoric correlations were statistically significant for both occupational specialties, the researcher believes the results lacked practical significance. It is conceivable that the significance resulted from the relatively large N (where N = students multiplied by the number of duty areas tested) rather than from any significant relationships between the two variables under investigation.

Recommendations for Further Research

Based on the findings of this investigation, tests developed using the Michigan State University vocational/occupational test-item bank seemed to possess reliability and validity. The item-bank method of developing vocational/occupational assessment tests should be investigated further with regard to reliability and validity. Tangential research topics include the following:

1. A similar investigation could be conducted for the remaining occupational specialties identified by V-TECS.

2. The present study could be replicated with a larger sample of vocational/occupational students and instructors.

3. The possibility of cross-walking all state vocational/occupational duty and task listings to develop national listings incorporating all common and geographically specific duties and tasks for each V-TECS occupational specialty could be studied.

Epilogue

For the past 18 months, the researcher has been involved in an investigation of vocational/occupational testing, specifically, tests developed at Michigan State University to assess vocational/occupational students in the areas of mechanical drafting and machine trades. Although the present study was focused on the reliability and validity of these tests, as an educator the researcher cannot overlook the people who are directly and indirectly affected by this assessment process.

With regard to assessment, two major observations stood out as paramount after the study was completed. First, a large percentage of instructors overrated and/or underrated their students' knowledge of the subject matter. In some cases, more than 50% of the students were ranked higher by their instructors than they actually achieved on the tests designed for them. The researcher has observed labs and classrooms designed primarily for the "performance" of job-related duties and tasks, not necessarily the cognitive development of the concepts related to those practices and procedures. As educators call for the rebirth of "problem-solving" skill development rather than learning by traditional methods, perhaps vocational instructors are placing too much emphasis on "projects" and psychomotor skill development and paying too little attention to the broad scope of knowledge (math/science) encompassed in each occupational specialty.

The second observations concerns the notion that a "core" curriculum currently exists within each occupational specialty. This belief has not been supported by research. Actually, a vast diversity exists in the course content and equipment used to deliver what is theoretically the same vocational program (program title) within the same state, and in some cases the same city. Even programs that appear to be similar (if not identical) because of equipment and local labor market needs are assessed in different areas with varying numbers of items. The content (curriculum) is determined by each individual instructor, not by a statewide committee or by common state curriculum guides.

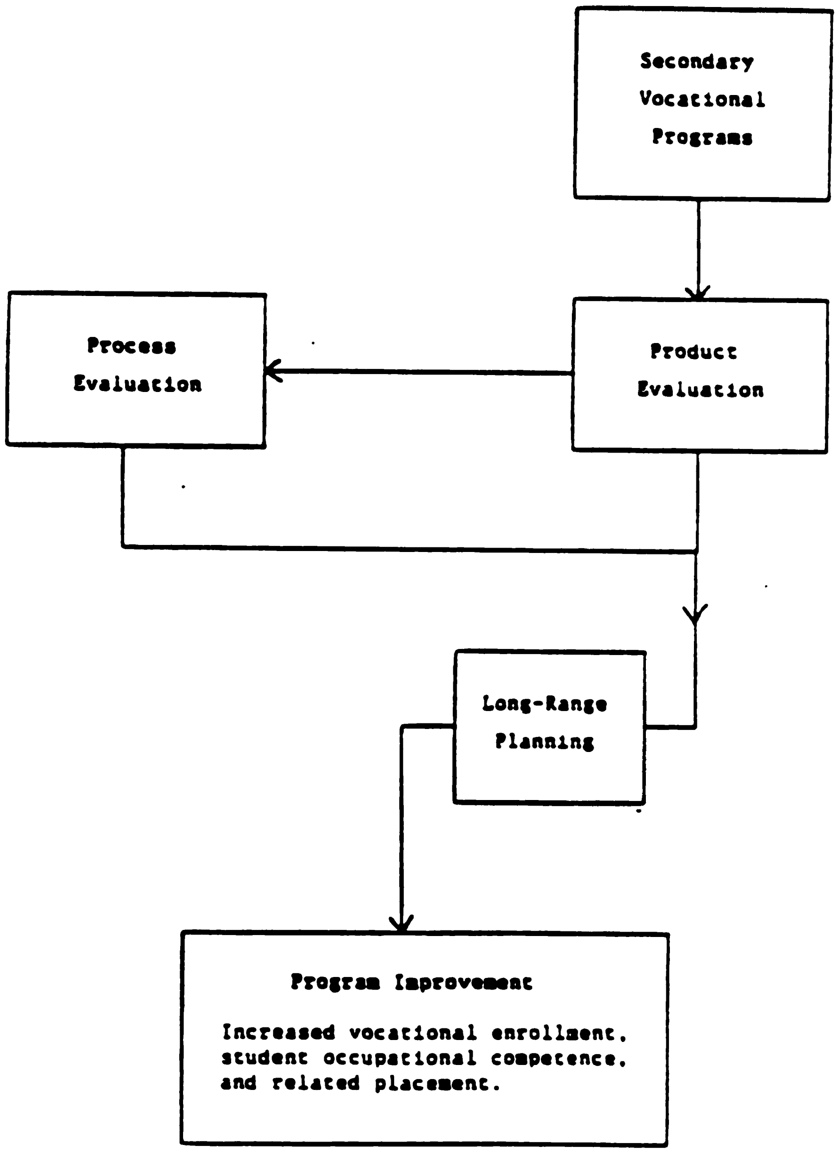
Finally, testing is an interaction between the instructor and the students he/she instructs. The result should determine, in prescriptive terms, the area(s) in which improvement is required so decisions can be made regarding instruction, curriculum, equipment, and so on. A test alone should <u>not</u> be used for decision making without supporting data, especially if the "cost" of making an incorrect decision is high.

APPENDICES

APPENDIX A

VOCATIONAL PROGRAM EVALUATION MODEL

Vocational Program Evaluation Model



APPENDIX B

MECHANICAL DRAFTING AND MACHINE TRADES TASK LISTINGS

FROM THE V-TECS

480100MV MECHANICAL DRAFTING (V) 02/01/85 406

SEE CIPS: 150202, 150805, 480105
DOT 007.281-010 - DRAFTER, MECHANICAL

THIS LISTING WAS DEVELOPED FOR THE VOCATIONAL-TECHNICAL EDUCATION CONSORTIUM OF STATES BY PERSONS IN THE STATE OF INDIANA.

DUTY AND TASK LISTING

A. DESIGNING ASSEMBLY TOOLING

1. DESIGN ASSEMBLY MACHINE
2. DESIGN AUTOMATIC PARTS HANDLING MACHINE
3. DESIGN CRIMPING DIE
4. DESIGN LOCATING FIXTURE
5. DESIGN WELDING FIXTURE
6. DESIGN STUD DRIVING FIXTURE

B. DESIGNING CUTTING DIES

1. DESIGN CUTOFF DIE
2. DESIGN TRIMMING DIE

C. DESIGNING GAGES

1. DESIGN FLUSH PIN GAGE
2. DESIGN PLUG GAGE
3. DESIGN PROFILE GAGE

D. DESIGNING MOLDS

1. DESIGN DIE CAST MOLD
2. DESIGN INJECTION MOLD

E. DESIGNING SPECIALTY FIXTURES

1. DESIGN GRINDING FIXTURE
2. DESIGN INSPECTION FIXTURES
3. DESIGN ABRASION TEST FIXTURE
4. DESIGN CHECKING (GAGE) FIXTURE
5. DESIGN COLD BOX FIXTURES
6. DESIGN ELECTRONIC CHECK FIXTURE
7. DESIGN LAPPING FIXTURE
8. DESIGN OVER HEAT FIXTURES
9. DESIGN REPAIR FIXTURES
10. DESIGN SONIC WELD FIXTURE

F. DETAILING SINGLE PARTS

1. DETAIL CAST MACHINING PART
2. DETAIL DEVELOPED PART
3. DETAIL DIE CAST PART
4. DETAIL EXTRUDED PART
5. DETAIL FORGING MACHINED PART
6. DETAIL FORGING PART
7. DETAIL LOST WAX CASTING PART
8. DETAIL LOW PRESSURE (VACUUM) CAST PART
9. DETAIL MACHINED PART
10. DETAIL PLASTIC MOLDED PART
11. DETAIL POWERED METAL PART
12. DETAIL PURCHASED PART
13. DETAIL REVISED PART
14. DETAIL SAND CASTING PART
15. DETAIL SHEET METAL PART
16. DETAIL TABULATED PART
17. DETAIL SINGLE PART
18. DETAIL PIPING
19. DETAIL CONTOURING TEMPLATE
20. DETAIL RUBBER MOLDED PART

G. DRAWING LAYOUTS

1. DRAW DETAIL ASSEMBLY
2. DRAW EXPERIMENTAL LAYOUT ASSEMBLY

3. DRAW EXPERIMENTAL LAYOUT SUB-ASSEMBLY
4. DRAW INSTALLATION ASSEMBLY
5. DRAW LAYOUT ASSEMBLY
6. DRAW LAYOUT SUB-ASSEMBLY
7. DRAW PART IDENTIFICATION ASSEMBLY
8. DRAW SPECIFICATION (OUTLINE) ASSEMBLY
9. DRAW SUB-ASSEMBLY
10. DRAW WORKING ASSEMBLY
11. DRAW COMPRESSOR ROOM LAYOUT

H. DRAWING CHARTS

1. DRAW BAR CHARTS
2. DRAW SECTOR CHART
3. DRAW SEQUENCE OF OPERATION CHART

I. DRAWING DIAGRAMS

1. DRAW ELECTRICAL SCHEMATIC DIAGRAM
2. DRAW FLOW DIAGRAM
3. DRAW HYDRAULIC SCHEMATIC DIAGRAM
4. DRAW PNEUMATIC SCHEMATIC DIAGRAM
5. DRAW WIRING DIAGRAM

J. DRAWING LAYOUTS OF SINGLE PARTS

1. DRAW CAST MACHINING PART
2. DRAW DEVELOPED PART
3. DRAW DIE CAST PART
4. DRAW FORGING MACHINED PART
5. DRAW MACHINED PART
6. DRAW PLASTIC MOLDED PART
7. DRAW PURCHASED PART
8. DRAW REVISED PART
9. DRAW SAND CASTING PART
10. DRAW SHEET METAL PART
11. DRAW CERAMIC PART
12. DRAW LAYOUT OF PRINTED CIRCUIT BOARD
13. DRAW RUBBER MOLDED PART

K. DRAWING MACHINE DIAGRAMS

1. DRAW ELECTRICAL SCHEMATIC DIAGRAM OF MACHINE

L. DRAWING PICTORIALS

1. DRAW ASSEMBLY PERSPECTIVE
2. DRAW SINGLE PART ISOMETRIC
3. DRAW SINGLE PART PERSPECTIVE

M. MAINTAINING AND CARING FOR TOOLS

1. ADJUST HEIGHT OF DRAFTING ARM OF ELBOW TYPE DRAFTING MACHINE
2. ADJUST HEIGHT OF DRAFTING ARM OF TRACK TYPE DRAFTING MACHINE
3. ADJUST SCALES
4. ALIGN INDEXING HEAD WITH BOARD
5. CLEAN BOARDS
6. CLEAN DRAFTING INSTRUMENTS
7. CLEAN SCALES
8. CLEAN TRIANGLES
9. EMPTY LEAD POINTER
10. LUBRICATE DRAFTING MACHINE
11. SET BASE LIVE ON TRACK TYPE MACHINE
12. ADJUST PARALLEL BAR
13. CLEAN RAPIDOGRAPH PEN
14. MAINTAIN BLUEPRINT FILE
15. MAINTAIN COMPUTER

N. REPRODUCING DRAWINGS

1. FILE PRINTS
2. MAKE MICROFILM APERTURE CARD OF DRAWING
3. MAKE MYLAR PRINT OF DRAWING
4. MAKE PRINT FROM APERTURE CARD

5. MAKE SEPIA PRINT OF DRAWINGS
6. RUN BLUEPRINT OF DRAWING
7. PRODUCE DRWING ON PLOTTER

O. TROUBLE-SHOOTING AND CORRECTING DESIGN ERRORS

1. TROUBLE-SHOOT CAUSE OF DIE COMPONENT BREAKAGE
2. TROUBLE-SHOOT CAUSE OF FAILURE OF PARTS TO BE REMOVED FROM THE DIE
3. TROUBLE-SHOOT CAUSE OF FAILURE OF PARTS TO BE REMOVED FROM THE FIXTURE
4. TROUBLE-SHOOT CAUSE OF FAILURE OF SCRAP TO BE REMOVED FROM THE DIE
5. TROUBLE-SHOOT CAUSE OF FIXTURE COMPONENT BREAKAGE
6. TROUBLE-SHOOT CAUSE OF IMPROPER FIXTURE CLEARANCES
7. TROUBLE-SHOOT CAUSE OF IMPROPER PART LOCATION IN FIXTURE
8. TROUBLE-SHOOT CAUSE OF INCORRECT PUNCH AND DIE CLEARANCE
9. TROUBLE-SHOOT CAUSE OF MISFITTING DIE COMPONENTS
10. TROUBLE-SHOOT CAUSE OF MISFITTING FIXTURE COMPONENTS
11. TROUBLE-SHOOT COMPUTER PROGRAM
12. TROUBLE-SHOOT MOLDING PROBLEMS

480500MV MACHINE SHOP (V) 02/01/85 451

SEE CIPS: 480503
DOT 600.280-022 - MACHINIST
O.E. CODE - 17.2302

THIS LISTING WAS DEVELOPED FOR THE VOCATIONAL-TECHNICAL EDUCATION CONSORTIUM OF STATES BY PERSONS IN THE STATE OF GEORGIA.

DUTY AND TASK LISTING

A. PERFORMING SUPERVISORY FUNCTIONS

1. COORDINATE WORKERS WITH WORK TO BE DONE
2. DETERMINE AVAILABILITY OF SUPPLIES AND MATERIALS
3. MAINTAIN FILES AND ORDER AND RECEIVE STOCK
4. FOLLOW-UP ON END PRODUCT QUALITY CONTROL STANDARDS
5. SUPERVISE MACHINE USE
6. SUPERVISE MAINTENANCE OF SHOP SAFETY EQUIPMENT, SAFE OPERATIONS AND SET UP OF MACHINERY
7. INVENTORY SUPPLIES AND MATERIALS

B. PERFORMING MATHEMATICAL CALCULATIONS

1. APPLY DIMENSIONS OF PARTS FROM BLUEPRINTS OR SPECIFICATIONS TO JOB
2. CALCULATE AMOUNT OF MATERIAL TO BE REMOVED TO OBTAIN CORRECT LIMITS FOR REWORK
3. CALCULATE CHANGE GEARING FOR THREADING
4. CALCULATE CONVERSION OF REVOLUTIONS PER MINUTE (RPM) TO SURFACE FEET PER MINUTE (SFPM)
5. CALCULATE DIMENSIONS OF SLOTS AND GROOVES ON SPECIAL SETUPS
6. CALCULATE GEAR BLANK SPECIFICATIONS FOR INDEXING
7. CALCULATE MACHINE SPEEDS AND FEEDS BY FORMULAS
8. CALCULATE STOCK UTILIZATION IN MACHINE WORK
9. CALCULATE TOLERANCES OR ALLOWANCES FOR PROPER FITS
10. CONVERT TO METRIC MEASUREMENT
11. DETERMINE CLEARANCE, RELIEF, AND RAKE OF CUTTING TOOLS
12. DETERMINE MATERIAL STRENGTH ACCORDING TO STANDARD RAW STOCK SIZES
13. TAKE MICROMETER READINGS
14. USE MACHINING HANDBOOKS, CHARTS, AND TABLES TO AID IN MATHEMATICAL CALCULATIONS

C. DESIGNING AND PLANNING MACHINE WORK

1. MAKE SKETCHES OF PARTS TO BE MACHINED
2. PERFORM LAYOUT FOR PRECISION MACHINE WORK USING LAYOUT INSTRUMENTS
3. PERFORM LAYOUT FOR PRECISION MACHINE WORK USING MILLING MACHINE
4. INSPECT, REMOVE, AND REPLACE PART(S) FOR REPAIR OR MACHINE WORK
5. REFER TO TECHNICAL ORDERS, STANDARDS, AND SPECIFICATIONS FOR MACHINE WORK
6. TEST FOR HARDNESS

D. PERFORMING METALWORK OPERATIONS

1. CLAMP WORK IN HOLDING DEVICE
2. CUT METAL STOCK
3. FABRICATE STOCKAL METALWORKING TOOLS
4. HEAT TREAT METAL
5. OPERATE CYLINDRICAL GRINDER
6. OPERATE HONE
7. PERFORM BENCH FILING
8. POLISH METAL
9. USING DEPTH GAGES
10. MEASURE WITH DIAL TEST INDICATOR
11. CHECK LAYOUT WITH GAGE BLOCKS
12. DETERMINE ACCURACY OF PRECISION MEASUREMENT INSTRUMENTS WITH GAGE BLOCKS
13. MEASURE WITH HEIGHT GAGES
14. MEASURE WITH SINE BAR

E. PERFORMING BENCH WORK

1. CUT MATERIALS WITH HAND HACKSAWS
2. CUT THREADS WITH HAND TAPS
3. CUT THREAD DIES
4. DISASSEMBLE AND ASSEMBLE PARTS
5. HONE AND LAP SURFACES

6. HAND SHARPEN CUTTING TOOLS WITH ABRASIVE STONES
7. REAM HOLES WITH HAND REAMERS
8. REMOVE AND REPLACE HELICAL COIL WIRE SCREW INSERT (STI)
9. REMOVE DAMAGED SCREWS AND OTHER NON-HARDENED THREADED HARDWARE
10. WORK AND SHAPE METAL

F. OPERATING DRILL PRESSES

1. CENTER PUNCH HOLE LOCATION
2. CONSIDER MATERIAL OF WORKPIECE TO BE DRILLED
3. COUNTERBORE TO DEPTH SPECIFIED IN BLUEPRINT
4. COUNTERSINK HOLE TO DRAWING REQUIREMENTS
5. DETERMINE HOLDING TECHNIQUE
6. DRILL HOLE TO SIZE
7. MOUNT AND SECURE WORK IN HOLDING DEVICE
8. REAM HOLE TO SIZE
9. SHARPEN DRILL USING GRINDING WHEEL
10. SHARPEN DRILL BIT FREE HAND AGAINST GRINDING WHEEL AND CHECK FOR SHARPNESS
11. SELECT TOOL OR CUTTER FOR DRILL PRESS OPERATION
12. SECURE TOOL OR CUTTER IN DRILL PRESS SPINDLE
13. SELECT SLEEVE TO SUIT DRILL
14. SET DRILL PRESS FOR PROPER FEEDS AND SPEEDS FOR OPERATION PERFORMED
15. SPOT-FACE WORKPIECE
16. USE TAPPING ATTACHMENT TO TAP HOLE
17. HAND TAP HOLE TO BLUEPRINT SPECIFICATIONS
18. USE AUTOMATIC FEED ON DRILL PRESS

G. OPERATING GRINDING MACHINES

1. ATTACH AND ALIGN MATERIALS FOR GRINDING OPERATION
2. BALANCE GRINDING WHEEL
3. CUT OFF OR PART MATERIALS WITH GRINDING MACHINES
4. DRESS AND TRUE GRINDING WHEELS
5. INSPECT GRINDING WHEELS
6. MEASURE, INSPECT, AND REWORK WORKPIECE ON GRINDING MACHINES
7. PERFORM GRINDING MACHINE OPERATIONS AS PER SETUP
8. POLISH WITH GRINDING MACHINE
9. SELECT AND SET SPEEDS AND FEEDS OF POWER FEED GRINDERS
10. SET UP AND PERFORM SURFACE GRINDING OPERATIONS
11. SET UP GRINDER AND SHARPEN PLAIN MILLING CUTTERS
12. SET UP, GRIND, AND SHARPEN PRESHAPED LATHE TOOLS
13. SET UP GRINDER AND SHAPE CHISELS
14. SET UP GRINDER TO RUN WORKPIECE BETWEEN CENTERS
15. SET UP GRINDER TO RUN ON MAGNETIC CHUCK
16. USE UTILITY GRINDER

H. SERVICING LATHES

1. ALIGN LATHE CENTERS USING APPROXIMATE METHOD
2. ALIGN LATHE CENTERS USING ACCURATE MEASUREMENT
3. BORE HOLES WITH LATHE
4. COUNTERBORE HOLES WITH LATHE
5. COUNTERSINK HOLES USING LATHE
6. USING TAPER ATTACHMENT, CUT LONG EXTERNAL TAPERED SURFACE
7. USING COMPOUND REST, CUT SHORT EXTERNAL TAPERED SURFACES
8. CUT INTERNAL THREADS WITH LATHE
9. CUT EXTERNAL THREADS WITH LATHE
10. CUT INTERNAL TAPERED SURFACES
11. DIE CUT THREADS WITH LATHE, HAND THREADING
12. DIE CUT THREADS WITH LATHE USING DIE HEADS
13. DRILL HOLES WITH LATHE
14. SET UP LATHE AND FACE WORKPIECE HELD IN A CHUCK
15. MEASURE STOCK
16. PERFORM CONTOUR, ANGULAR, OR RADII CUTS WITH LATHE
17. PERFORM LATHE FILING
18. PERFORM LATHE FILING TO DEBURR PART
19. PERFORM SPINNING OPERATION USING FORMING TOOL
20. REAM HOLES WITH LATHE
21. RECHASE THREADS ON LATHE
22. ROUGH CUT AND FINISH CUT WITH LATHE
23. KNURL PARTS WITH LATHE
24. SECURE TOOL HOLDER, FIXTURES, OR ATTACHMENTS
25. SELECT AND SET SPEEDS AND FEEDS

26. SET UP ENGINE LATHE
27. SET UP TURRET LATHE FOR OPERATIONS
28. TAP THREADS WITH LATHE
29. SET UP TOOL POST GRINDER
30. PERFORM GRINDING OPERATION

I. OPERATING MILLING MACHINES

1. ALIGN MILLING MACHINE FIXTURES
2. ALIGN MILLING MACHINE ATTACHMENTS
3. ASSEMBLE COMPLETED MILL WORK
4. BORE HOLES WITH MILLING MACHINES
5. BORE TO FINISH BUSHING
6. BORE TO REMOVE BUSHINGS
7. CUT EXTERNAL KEYWAY
8. DRILL HOLES WITH MILLING MACHINE
9. DUPLICATE ON THE PROFILE MILLING MACHINE
10. INSPECT COMPLETED MILL WORK
11. MILL AN ANGLE
12. MILL AN EXTERNAL RADIUS
13. MILL CYLINDRICAL WORK
14. MILL GEARS
15. MILL INTERNAL SLOTS USING SLOTTER AND ATTACHMENTS
16. PERFORM END MILLING
17. PERFORM FLYCUTTING OPERATIONS
18. PERFORM FORM MILLING
19. PERFORM INDEXING OPERATIONS
20. PERFORM REAMING OPERATIONS
21. PERFORM CUTTING-OFF OPERATIONS
22. PERFORM STRADDLMOOOLLING OPERATIONS ON THE HORIZONTAL MILL
23. SELECT AND SET SPEEDS AND FEEDS FOR MILLING WORK
24. SQUARE UP METAL USING DIVIDING HEAD
25. SQUARE UP METAL USING TABLE VISE

J. OPERATING POWER SAWS

1. CUT AND WELD BANDSAW BLADES TO INSERT FOR CONTOUR SAWING
2. MEASURE MATERIAL AND CUT OFF MATERIAL WITH POWER HACK SAW
3. DETERMINE TYPE OF MATERIAL TO BE SAWED
4. REMOVE AND REPLACE SAW BLADES
5. SAWING WITH A METAL BAND SAW TO SCRIBED LINES
6. SELECT AND SET SPEEDS AND FEEDS FOR SAWING OPERATIONS
7. SELECT APPROPRIATE BLADES

K. OPERATING PRESSES

1. SET UP AND PUNCH MATERIALS WITH PRESS
2. SELECT ACCESSORIES AND ATTACHMENTS FOR PRESS WORK
3. SET UP PRESS AND ASSEMBLE PARTS WITH PRESSES
4. SET UP PRESS AND DISASSEMBLE PARTS WITH PRESSES
5. STRAIGHTEN MISCELLANEOUS PARTS WITH PRESSES
6. SET UP, REMOVE, AND REPLACE PARTS WITH PRESSES

L. OPERATING SHAPERS

1. ALIGN SHAPER ATTACHMENT, WORKPIECE, AND CUTTING TOOL
2. SELECT AND SET SPEEDS AND FEEDS OF SHAPERS
3. SELECT, SHAPE, AND SHARPEN CUTTING TOOLS FOR SHAPER OPERATION

M. PERFORMING PRODUCTION MACHINIST LINE

1. REMOVE AND INSTALL PINS
2. REMOVE DEFECTIVE SPLINES AND REPLACE WITH NEW SPLINES
3. REMOVE FROZEN OR SEIZED PARTS

N. MAINTAINING MACHINES AND TOOLS

1. INSPECT AND CHANGE DRIVE PULLEYS OR BELTS
2. CLEAN AND STORE HAND TOOLS, CUTTERS, FIXTURES, JIGS, OR ATTACHMENTS
3. INSTALL, LEVEL, AND FASTEN DOWN MACHINES
4. REMOVE MACHINE PARTS
5. REPLACE AND ADJUST MACHINE PARTS
6. INSPECT AND REMOVE, REPLACE, OR ADJUST MACHINE GUARDS
7. SCRAPE AND PAINT MACHINES

8. INSPECT AND REPAIR HAND TOOLS
9. STORE GRINDING WHEELS
10. STORE PRECISION TOOLS
11. PERFORM MAINTENANCE ON LATHE
12. PERFORM MAINTENANCE ON MILLING MACHINE
13. PERFORM MAINTENANCE ON DRILL PRESS
14. PERFORM MAINTENANCE ON GRINDER
15. PERFORM MAINTENANCE ON BANDSAW
16. SELECT COOLANTS, CUTTING OILS, OR COMPOUNDS FOR MACHINING OPERATION
17. CLEAN ARBOR PRESSES
18. LUBRICATE ARBOR PRESS
19. CLEAN HYDRAULIC PRESS
20. LUBRICATE HYDRAULIC PRESSES
21. INSPECT ARBOR PRESSES FOR SAFE OPERATIONAL CONDITION
22. INSPECT HYDRAULIC PRESS FOR SAFE OPERATIONAL CONDITION
23. INSPECT WORK AREA FOR SAFE WORKING ENVIRONMENT

O. MAINTAINING SHOP FACILITIES AND WORK AREAS

1. DISPOSE OF SCRAP METAL, CHIPS OR SHAVINGS, AND TRASH OR WASTE MATERIALS
2. PERFORM CUSTODIAL TASKS
3. CLEAN AND MAINTAIN WORK AND ADJACENT WORK AREAS
4. PAINT FACILITIES
5. SWEEP AND CLEAN SHOP FACILITY

APPENDIX C

SAMPLE MECHANICAL DRAFTING AND MACHINE TRADES

TEST BLUEPRINTS

INSTRUCTOR'S NAME ___________________________________
SCHOOL/CENTER ___________________________________
ADDRESS ___________________________________
CITY/ZIP ___________________________________

TEST DATE: _____________
ESTIMATED DATE TEST WILL
BE RECEIVED IN RESOURCE OFFICE AT MSU: _____________

INSTRUCTIONS: REQUEST TEST ITEMS AT THE FULL TASK LEVEL, DUTY LEVEL, OR INDIVIDUAL TASK LEVEL BY INSERTING THE REQUESTED NUMBER OF ITEMS IN THE CORRESPONDING BOX. FOR HELP, SEE ATTACHED EXAMPLE.

☐ TASKLIST ID: 480100MV MECHANICAL DRAFTING (Copyright - 3-16-84)

☐ A. DESIGNING ASSEMBLY TOOLING

☐ 1. Design Assembly machine

☐ 2. Design automatic parts handling machine

☐ 3. Design Crimping die

☐ 4. Design locating fixture

☐ 5. Design welding fixture

☐ 6. Design stud driving fixture

☐ TOTAL FOR DUTY: A

☐ B. DESIGNING CUTTING DIES

☐ 1. Design cutoff die

☐ 2. Design trimming die

☐ TOTAL FOR DUTY: B

☐ C. DESIGNING GAGES

☐ 1. Design flush pin gage

☐ 9. Detail machined part

☐ 10. Detail plastic molded part

☐ 11. Detail powered metal part

☐ 12. Detail purchased part

☐ 13. Detail revised part

☐ 14. Detail sand casting part

☐ 15. Detail sheet metal part

☐ 16. Detail tabulated part

☐ 17. Detail single part

☐ 18. Detail piping

☐ 19. Detail contouring template

☐ 20. Detail rubber molded part

☐ TOTAL FOR DUTY: F

☐ G. DRAWING LAYOUTS

☐ 1. Draw detail assembly

☐ 2. Draw experimental layout assembly

☐ 3. Draw experimental layout subassembly

☐ 4. Draw installation assembly

☐ 5. Draw layout assembly

☐ 6. Draw layout subassembly

☐ 7. Draw part identification assembly

☐ 8. Draw specification (outline) assembly

☐ 9. Draw subassembly

☐ 10. Draw working assembly

☐ 11. Draw compressor room layout

☐ TOTAL FOR DUTY: G

☐ H. DRAWING CHARTS

☐ 1. Draw bar charts

☐ 2. Draw sector chart

INSTRUCTOR'S NAME ______________________
SCHOOL/CENTER ______________________
ADDRESS ______________________
CITY/ZIP ______________________

TEST DATE: __________
ESTIMATED DATE TEST WILL
BE RECEIVED IN RESOURCE OFFICE AT MSU: __________

INSTRUCTIONS: REQUEST TEST ITEMS AT THE FULL TASK LEVEL, DUTY LEVEL, OR INDIVIDUAL TASK LEVEL BY INSERTING THE REQUESTED NUMBER OF ITEMS IN THE CORRESPONDING BOX. FOR HELP, SEE ATTACHED EXAMPLE.

☐ TASKLIST ID: 480500MV MACHINE SHOP (Copyright - 12-16-76)

☐ A. PERFORMING SUPERVISORY FUNCTIONS

☐ 1. Coordinate workers with work to be done

☐ 2. Determine availability of supplies and materials

☐ 3. Maintain files and order and receive stock

☐ 4. Follow-up on end product quality control standards

☐ 5. Supervise machine use

☐ 6. Supervise maintenance of shop safety equipment, safe operations and set up of machinery

☐ 7. Inventory supplies and materials

☐ TOTAL FOR DUTY: A

☐ B. PERFORMING MATHEMATICAL CALCULATIONS

☐ 1. Apply dimensions of parts from blueprints or specifications to job

☐ 2. Calculate amount of material to be removed to obtain correct limits for rework

☐ 3. Calculate change gearing for threading

☐ 4. Calculate conversion of revolutions per minute (RPM) to surface feet per minute (SFPM)

☐ 5. Calculate dimensions of slots and grooves on special setups

☐ 6. Calculate gear blank specifications for indexing

☐ 7. Calculate machine speeds and feeds by formulas

☐ 8. Calculate stock utilization in machine work

☐ 9. Calculate tolerances or allowances for proper fits

☐ 10. Convert to metric measurement

☐ 11. Determine clearance, relief, and rake of cutting tools

☐ 12. Determine material strength according to standard raw stock sizes

☐ 13. Take micrometer readings

☐ 14. Use machining handbooks, charts, and tables to aid in mathematical calculations

☐ TOTAL FOR DUTY: B

☐ C. DESIGNING AND PLANNING MACHINE WORK

☐ 1. Make sketches of parts to be machined

☐ 2. Perform layout for precision machine work using layout instruments

☐ 3. Perform layout for precision machine work using milling machine

☐ 4. Inspect, remove, and replace part(s) for repair or machine work

☐ 5. Refer to technical orders, standards, and specifications for machine wor

☐ 6. Test for hardness

☐ TOTAL FOR DUTY: C

☐ D. PERFORMING METALWORK OPERATIONS

☐ 1. Clamp work in holding device

☐ 2. Cut metal stock

☐ 3. Fabricate special metalworking tools

☐ 4. Heat treat metal

☐ 5. Operate cylindrical grindeR

☐ 6. Operate hone

☐ 7. Perform bench filing

☐ 8. Polish metal

- [ ] 9. Using depth gages
- [ ] 10. Measure with dial test indicator
- [ ] 11. Check layout with gage blocks
- [ ] 12. Determine accuracy of precision measurement instruments with gage blocks
- [ ] 13. Measure with height gages
- [ ] 14. Measure with sine bar
- [ ] TOTAL FOR DUTY: D

- [ ] E. PERFORMING BENCH WORK
- [ ] 1. Cut materials with hand hacksaws
- [ ] 2. Cut threads with hand taps
- [ ] 3. Cut thread dies
- [ ] 4. Disassemble and assemble parts
- [ ] 5. Hone and lap surfaces
- [ ] 6. Hand sharpen cutting tools with abrasive stones
- [ ] 7. Ream holes with hand reamers
- [ ] 8. Remove and replace helical coil wire screw insert (STI)
- [ ] 9. Remove damaged screws and other nonhardened threaded hardware
- [ ] 10. Work and shape metal
- [ ] TOTAL FOR DUTY: E

- [ ] F. OPERATING DRILL PRESSES
- [ ] 1. Center punch hole location
- [ ] 2. Consider material of workpiece to be drilled
- [ ] 3. Counterbore to depth specified in blueprint
- [ ] 4. Countersink hole to drawing requirements
- [ ] 5. Determine holding technique
- [ ] 6. Drill hole to size
- [ ] 7. Mount and secure work in holding device
- [ ] 8. Ream hole to size
- [ ] 9. Sharpen drill using grinding wheel

APPENDIX D

CORRESPONDENCE

MEMORANDUM

TO: Student Assessment Advisory Committee/Pilot Site Contacts

FROM: Christan Olson and Steven C. Clark

RE: Core Curriculum

DATE: February 18, 1986

Some of you have expressed an interest in the capability of comparing your student assessment test scores with those of similar students throughout the state. The primary problem is that there is no guarantee that all students have studied the same material and developed competencies in the same tasks. As you know, Michigan does not have a 'core curriculum' but rather curriculum guides that allow teachers to choose the tasks to be included in the program with input from local resources.

In an attempt to assist those that would like to compare their students with others in like program areas, we would like to survey participating teachers to identify the tasks taught in their program. The Curriculum Resource Team will compare responses to identify those tasks taught in every program.

To assist us, would you please identify a teacher (or teachers) in each of the following program areas who will be nominated to participate in the 'student assessment pilot test' in April, May and June of this year. We will then send them a task list and request that they 'highlight' the tasks taught in their programs. We will then be able to compare the results and report back to you at a later date. We would like to have the responses back from teachers by the end of February.

You do not need a person in each category, only those who you wish to participate in the pilot testing.

Thank you for your assistance.

Nominations for Participation in Student Assessment Pilot Test

Program	Teacher Name (Address if different than yours)
Cashier/Checker	____
Cosmetology	____
Data Processing	____
Diesel Mechanics	____
Electronics	____
Food Preparation	____
Horticulture	____
Machine Trades	____
Mech. Drafting	____
Nurse Assisting	____
Welding	____

Return to:

Chris M. Olson, Ph.D.
102 Wills House
Michigan State University
East Lansing, MI 48824

-OR-

Steven Clark, Research Assistant
102 Wills House
Michigan State University
East Lansing, MI 48824

Phone (in Michigan): 1-800-831-7523

MICHIGAN STATE UNIVERSITY
CURRICULUM RESOURCE TEAM

DEPARTMENT OF AGRICULTURAL AND EXTENSION EDUCATION
COLLEGE OF AGRICULTURE AND NATURAL RESOURCES
TELEPHONE(517) 353-0661

March 4, 1986

Dear

You have been identified as a potential participant in a student assessment pilot test, by . Your participation will depend on the availability of test items, and you will be contacted concerning this later.

A first step in the process, however, requires us to identify any commonalities among programs in different areas of the state. Would you please highlight on the attached task list(s), those tasks taught in your program. Your response by March 17 will be appreciated.

Sincerely,

Chris M. Olson, Ph.D.
Project Director and
Associate Professor

cc: Mack Seney
Steve Clark

MSU Curriculum Resource Team
101 Wills House
Michigan State University
East Lansing, Michigan 48824-1050

MICHIGAN STATE UNIVERSITY
CURRICULUM RESOURCE TEAM

DEPARTMENT OF AGRICULTURAL AND EXTENSION EDUCATION
COLLEGE OF AGRICULTURE AND NATURAL RESOURCES

TELEPHONE(517) 353-0661

April 22, 1986

TO: Pilot Site Instructors

FROM: Steven Clark, Research Assistant

RE: Development of Test Form A1/B1 (Directions)

Enclosed please find a copy of the V-TECS duty and task list for your program. As discussed during the meeting last month, you are to choose from the task listing, the duties and tasks that most correctly parallel your curriculum content. A test will be developed based on this information (Form A1 and parallel Form B1) and sent to you. Please include the number of items you wish for each duty area. I will include an additional 25% for each area, than you will be able to select only those items which you want to see on your final test (Form A2 and B2).

Form A1 and B1 come in two parts; (1) the test questions and, (2) the answer key. Form A1/B1 will contain at least 25% more items than you wish on you final test (Form A2/B2).

1. Read each of the test questions carefully. Determine whether on not the content of the question reflects content taught in your curriculum.
2. Check the answer key carefully. Make sure the answer is the correct one.
3. On the answer key, circle those questions YOU WISH TO DELETE. This is local selection (validation) of test content. You will have a test that reflects you local curriculum and is content valid.
4. If you should delete more items than the 25%, additional items may be added. Contact me as soon as possible.
5. Return the answer key (with the questions circled that you wish deleted) to me for modification into Forms A2 (test) and B2 (retest).
6. Indicate the number of students that will be tested so I will have enough answer sheets with me on test day.

Thank-you. Call if you have any questions. I will see you on your scheduled test day.

MSU Curriculum Resource Team
101 Wills House
Michigan State University
East Lansing, Michigan 48824-1050

MICHIGAN STATE UNIVERSITY

UNIVERSITY COMMITTEE ON RESEARCH INVOLVING HUMAN SUBJECTS (UCRIHS)
238 ADMINISTRATION BUILDING
(517) 355-2186

EAST LANSING • MICHIGAN • 48824-1046

April 23, 1986

Mr. Steven C. Clark
Counseling, Educational Psychology
and Special Education
Erickson Hall

Dear Mr. Clark:

Subject: Proposal Entitled, "Testing the Reliability and Validity of Occupational Competency Assessment Instruments Constructed Via Random Selection of Previously Validated Test Items"

I am pleased to advise that I concur with your evaluation that this project is exempt from full UCRIHS review, and approval is herewith granted for conduct of the project.

You are reminded that UCRIHS approval is valid for one calendar year. If you plan to continue this project beyond one year, please make provisions for obtaining appropriate UCRIHS approval prior to April 23, 1987.

Any changes in procedures involving human subjects must be reviewed by the UCRIHS prior to initiation of the change. UCRIHS must also be notified promptly of any problems (unexpected side effects, complaints, etc.) involving human subjects during the course of the work.

Thank you for bringing this project to my attention. If I can be of any future help, please do not hesitate to let me know.

Sincerely,

Henry E. Bredeck
Chairman, UCRIHS

HEB/jms

cc: Dr. Chris M. Olson

APPENDIX E

SAMPLE INSTRUCTOR RANKING SHEETS FOR MECHANICAL DRAFTING AND MACHINE TRADES

PILOT TESTING
STUDENT ASSESSMENT
INSTRUCTOR PREDICTION SHEETS

Student Number: Instructor Code:

School Code: Program Code: 480100MV

INSTRUCTIONS: Please check the skills/knowledge possessed by the student (either Pass of Fail based on your subjective judgement) using the criterion 50% and above for a Pass, and 49% and below for a Fail.

EXAMPLE OF A PASS: If questioned, "sample student" possesses enough skills/knowledge to successfully achieve a rating of 50% or higher within the stated duty area.

EXAMPLE OF A FAIL: If questioned, "sample student" does not possess enough skills/knowledge to successfully achieve a rating of 50% or higher within the stated duty area.

DUTY AREAS

Pass ☐ Fail ☐

F. DETAILING SINGLE PARTS
1. Detail cast machining part
2. Detail developed part
3. Detail die cast part
4. Detail extruded part
5. Detail forging machined part
6. Detail forging part
7. Detail lost wax casting part
8. Detail low pressure (vacuum) cast part
9. Detail machined part
10. Detail plastic molded part
11. Detail powered metal part
12. Detail purchased part
13. Detail revised part
14. Detail sand casting part
15. Detail sheet metal part
16. Detail tabulated part
17. Detail single part
18. Detail piping
19. Detail contouring template
20. Detail rubber molded part

Pass ☐ Fail ☐

J. DRAWING LAYOUTS OF SINGLE PARTS
1. Draw cast machining part
2. Draw developed part
3. Draw die cast part
4. Draw forging machined part
5. Draw machined part
6. Draw plastic molded part
7. Draw purchased part
8. Draw revised part
9. Draw sand casting part
10. Draw sheet metal part
11. Draw ceramic part
12. Draw layout of printed circuit board
13. Draw rubber molded part

DUTY AREAS

Pass ☐ Fail ☐

G. DRAWING LAYOUTS
1. Draw detail assembly
2. Draw experimental layout assembly
3. Draw experimental layout subassembly
4. Draw installation assembly
5. Draw layout assembly
6. Draw layout subassembly
7. Draw part identification assembly
8. Draw specification (outline) assembly
9. Draw subassembly
10. Draw working assembly
11. Draw compressor room layout

Pass ☐ Fail ☐

H. DRAWING CHARTS
1. Draw bar charts
2. Draw sector chart
3. Draw sequence of operation chart

Pass ☐ Fail ☐

I. DRAWING DIAGRAMS
1. Draw electrical schematic diagram
2. Draw flow diagram
3. Draw hydraulic schematic diagram
4. Draw pneumatic schematic diagram
5. Draw wiring diagram

DUTY AREAS

Pass ☐ Fail ☐

L. DRAWING PICTORIALS
1. Draw assembly perspective
2. Draw single part isometric
3. Draw single part perspective

Pass ☐ Fail ☐

N. REPRODUCING DRAWINGS
1. File Prints
2. Make microfilm aperture card of drawing
3. Make mylar print of drawing
4. Make print from aperture card
5. Make sepia print of drawings
6. Run blueprint of drawing
7. Produce drawing on plotter

DUTY AREAS

Pass ☐ Fail ☐

M. MAINTAINING AND CARING FOR TOOLS
1. Adjust height of drafting arm of elbow type drafting machine
2. Adjust height of drafting arm of track type drafting machine
3. Adjust scales
4. Align indexing head with board
5. Clean boards
6. Clean drafting instruments
7. Clean scales
8. Clean triangles
9. Empty lead pointer
10. Lubricate drafting machine
11. Set base live on track type machine
12. Adjust parallel bar
13. Clean rapidograph pen
14. Maintain blueprint file
15. Maintain computer

* Please include this students final letter grade in your course _________.

* Because letter grades are, in many cases, not based totally on acquisition of knowledge/skills, based on your subjective judgement check the box (percentage group) in which you feel this student would fall based solely on knowledge/skill level.

%	0 - 19	20 - 39	40 - 59	60 - 79	80 - 100
	☐	☐	☐	☐	☐

Thank you for your time and effort.

PILOT TESTING
STUDENT ASSESSMENT
INSTRUCTOR PREDICTION SHEETS

Student Number: Instructor Code:

School Code: Program Code: 480500MV

INSTRUCTIONS: Please check the skills/knowledge possessed by the student (either Pass or Fail based on your subjective judgement) using the criterion 50% and above for a Pass, and 49% and below for a Fail.

EXAMPLE OF A PASS: If questioned, "sample student" possesses enough skills/knowledge to successfully achieve a rating of 50% or higher within stated duty area.

EXAMPLE OF A FAIL: If questioned, "sample student" does not possess enough skills/knowledge to successfully achieve a rating of 50% or higher within the stated duty area.

DUTY AREAS

Pass	Fail

A. PERFORMING SUPERVISORY FUNCTIONS
1. Coordinate workers with work to be done
2. Determine availability of supplies and materials
3. Maintain files and order and receive stock
4. Follow-up on end product quality control standards
5. Supervise machine use
6. Supervise maintenance of shop safety equipment, safe operations and set up of machinery
7. Inventory supplies and materials

Pass	Fail

C. DESIGNING AND PLANNING MACHINE WORK
1. Make sketches of parts to be machined
2. Perform layout for precision machine work using layout instruments
3. Perform layout for precision machine work using milling machine
4. Inspect, remove, and replace part(s) for repair or machine work
5. Refer to technical orders, standards, and specifications for machine work
6. Test for hardness

Pass	Fail

E. PERFORMING BENCH WORK
1. Cut materials with hand hacksaws
2. Cut threads with hand taps.
3. Cut threads with hand dies
4. Disassemble and assemble parts
5. Hone and lap surfaces
6. Hand sharpen cutting tools with abrasive stones
7. Ream holes with hand reamers
8. Remove and replace helical coil wire screw insert
9. Remove damaged screws and other nonhardened threaded hardware
10. Work and shape metal

DUTY AREAS

Pass	Fail

B. PERFORMING MATHEMATICAL CALCULATIONS
1. Apply dimensions of parts from blueprints or specifications to job sheets
2. Calculate change gearing for threading
3. Calculate amount of material to be removed to obtain correct limits for rework
4. Calculate conversion of revolutions per minute (RPM) to surface feet per minute (SFPM)
5. Calculate dimensions of slots and grooves on special setups
6. Calculate gear blank specifications for indexing
7. Calculate machine speeds and feeds by formulas
8. Calculate stock utilization in machine work
9. Calculate tolerances of allowances for proper fits
10. Convert to metric measurement
11. Determine clearance, relief, and rake of cutting tools
12. Determine material strength according to standard raw stock sizes
13. Take micrometer readings
14. Use machining handbooks, charts, and tables to aid in mathematical calculations

Pass	Fail

D. PERFORMING METALWORK OPERATIONS
1. Clamp work in holding device
2. Cut metal stock
3. Fabricate special metalworking tools
4. Heat treat metal
5. Operate cylindrical grinder
6. Operate hone
7. Perform bench filing
8. Polish metal
9. Use depth gages
10. Measure with dial test indicator
11. Check layout with gage blocks
12. Determine accuracy of precision measurement instruments with gage blocks
13. Measure with height gages
14. Measure with sine bar

Pass | Fail

F. OPERATING DRILL PRESSES

1. Center punch hole location
2. Consider material of workpiece to be drilled
3. Counterbore to depth specified in blueprint
4. Countersink hole to drawing requirements
5. Determine holding technique
6. Drill hole to size
7. Mount and secure work in holding device
8. Ream hole to size
9. Sharpen drill using grinding wheel
10. Sharpen drill bit free hand against grinding wheel and check for sharpness
11. Select tool or cutter for drill press operation
12. Secure tool or cutter in drill press spindle
13. Select sleeve to suit drill
14. Set drill press for proper feeds and speeds for operation performed
15. Spot-face workpiece
16. Use tapping attachment to tap hole
17. Hand tap hole to blueprint specifications
18. Use automatic feed on drill press

Pass | Fail

H. LATHE OPERATIONS

1. Align lathe centers using approximate method
2. Align lathe centers using accurate measurement
3. Bore holes with the lathe
4. Counterbore holes with lathe
5. Countersink holes using lathe
6. Using taper attachment, cut long external tapered surfaces
7. Using compound rest, cut short external tapered surfaces
8. Cut internal threads with lathe
9. Cut external threads with lathe
10. Cut internal tapered surfaces
11. Die cut threads with lathe, hand threading
12. Die cut threads with lathe using die heads
13. Drill holes with the lathe
14. Set up lathe, and face workpiece held in a chuck
15. Measure stock
16. Perform contour, angular, or radii cuts with lathe
17. Perform lathe filing to deburr part
18. Perform lathe filing
19. Perform spinning operation using forming tool
20. Ream holes with the lathe
21. Rechase threads on lathe
22. Rough cut and finish cut with lathe
23. Knurl parts with lathe
24. Secure tool holder, fixtures, or attachments
25. Select and set feeds and speeds
26. Set up engine lathe
27. Set up turret lathe for operations
28. Tap threads with lathe
29. Set up tool post grinder
30. Perform grinding operations

Pass | Fail

G. OPERATING GRINDING MACHINES

1. Attach and align materials for grinding operations
2. Balance grinding wheel
3. Cut off or part materials with grinding machines
4. Dress and true grinding wheels
5. Inspect grinding wheels
6. Measure, inspect, and rework workpiece on grinding machines
7. Perform grinding machine operations as per setup
8. Polish with grinding machine
9. Select and set speeds and feeds of power feed grinders
10. Set up grinder and sharpen plain milling cutters
11. Set up and perform surface grinding operations
12. Set up , grind, and sharpen preshaped lathe tools
13. Set up grinder and shape chisels
14. Set up grinder to run workpiece between centers
15. Set up grinder to run on magnetic chuck
16. Use utility grinder

Pass | Fail

I. OPERATING MILLING MACHINES

1. Align milling machine fixtures
2. Align milling machine attachments
3. Assemble completed mill work
4. Bore holes with milling machines
5. Bore to finish bushing
6. Bore to remove bushings
7. Cut external keyway
8. Drill holes with milling machines
9. Duplicate on the profile milling machine
10. Inspect completed mill work
11. Mill an angle
12. Mill an external radius
13. Mill cylindrical work
14. Mill gears
15. Mill internal slots using slotter and attachments
16. Perform end milling
17. Perform flycutting operations
18. Perform form milling
19. Perform indexing operations
20. Perform reaming operations
21. Perform cutting-off operations
22. Perform straddle milling operations on the on the horizontal milling machine
23. Select and set speeds and feeds for milling work
24. Square up metal using dividing head
25. Square up metal using mill table vise

Pass	Fail	
☐	☐	**J. OPERATING POWER SAWS**

1. Cut and weld bandsaw blades to insert for contour sawing
2. Measure material and cut off material with power hack saw
3. Determine type of material to be sawed
4. Remove and replace saw blades
5. Sawing with a metal band saw to scribed lines
6. Select and set speeds and feeds for sawing operations
7. Select appropriate blades

Pass	Fail	
☐	☐	**S. MAINTAINING SAFE SHOP FACILITIES AND WORK AREAS**

1. Dispose of scrap metal, chips or shavings, and trash or waste materials
2. Perform custodial tasks
3. Clean and maintain work and adjacent work areas
4. Paint facilities
5. Sweep and clean shop facility

* Please include this students final letter grade in your course ________.

* Because letter grades are in many cases not based totally on acquisition of knowledge/skills, based on your subjective judgement check the box (percentile group) in which you feel this student would fall based solely on knowledge/skill level.

%	0 - 19	20 - 39	40 - 59	60 - 79	80 - 100
	☐	☐	☐	☐	☐

Thank you for your time and effort.

APPENDIX F

DATA SUMMARY TABLES

Table F.1.--Vocational/occupational assessment test results, by site: machine trades, test Form A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Site A	50 - 60	1	10	100
	40 - 49	4	9	90
	30 - 39	3	5	70
	20 - 29	2	2	20
Site B	80 - 90	4	37	100
	70 - 79	0	33	89
	60 - 69	2	33	89
	50 - 59	0	31	84
	40 - 49	4	31	84
	30 - 39	12	27	73
	20 - 29	15	15	41
Site C	60 - 70	3	25	100
	50 - 59	4	22	88
	40 - 49	8	18	72
	30 - 39	3	3	12
Site D	60 - 70	2	21	100
	50 - 59	8	19	90
	40 - 49	10	11	52
	30 - 39	1	1	5
Site E	70 - 80	1	15	100
	60 - 69	3	14	93
	50 - 59	0	11	73
	40 - 49	7	11	73
	30 - 39	3	4	27
	20 - 29	1	1	6
Site F	60 - 70	8	22	100
	50 - 59	12	14	64
	40 - 49	2	2	9
Site G	50 - 60	1	3	100
	40 - 49	2	2	67

Table F.2.--Vocational/occupational assessment test results, by site: machine trades, retest From B2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Site A	50 - 60	1	5	100
	40 - 49	2	4	80
	30 - 39	2	2	40
Site B	50 - 60	1	21	100
	40 - 49	1	20	95
	30 - 39	5	19	90
	20 - 29	13	14	67
	10 - 19	1	1	5
Site C	50 - 60	4	12	100
	40 - 49	2	4	33
	20 - 29	2	2	17
Site D	60 - 70	2	11	100
	50 - 59	4	9	82
	40 - 49	4	5	45
	30 - 39	1	1	9
Site E	60 - 70	2	8	100
	50 - 59	1	6	75
	40 - 49	1	5	63
	30 - 39	1	4	50
	20 - 29	2	3	38
	10 - 19	1	1	13
Site F	70 - 80	1	13	100
	60 - 69	4	12	92
	50 - 59	6	8	62
	40 - 49	1	2	15
	30 - 39	1	1	8
Site G	60 - 70	1	2	100
	50 - 59	0	1	50
	40 - 49	0	1	50
	30 - 39	1	1	50

Table F.3.--Vocational/occupational assessment test results, by site: machine trades, retest Form A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Site A	40 - 50	4	5	100
	30 - 39	1	1	20
Site B	80 - 90	2	16	100
	70 - 79	0	14	88
	60 - 69	0	14	88
	50 - 59	0	14	88
	40 - 49	2	14	88
	30 - 39	1	12	75
	20 - 29	11	11	69
Site C	50 - 60	5	13	100
	40 - 49	4	8	62
	30 - 39	4	4	31
Site D	60 - 70	1	10	100
	50 - 59	1	9	90
	40 - 49	7	8	80
	30 - 39	1	1	10
Site E	50 - 60	3	7	100
	40 - 49	1	4	57
	30 - 39	3	3	43
Site F	60 - 70	6	9	100
	50 - 59	2	3	33
	40 - 49	1	1	11
Site G	40 - 50	1	1	100

Table F.4.—Machine trades summary, all sites: test Form A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Sites	80 - 90	4	133	100
A - G	70 - 79	1	129	97
	60 - 69	18	128	96
	50 - 59	26	110	83
	40 - 49	37	84	63
	30 - 39	26	47	35
	20 - 29	21	21	16

Table F.5.—Machine trades summary, all sites: retest Form A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Sites	80 - 90	2	61	100
A - G	70 - 79	0	59	97
	60 - 69	7	59	97
	50 - 59	11	52	85
	40 - 49	20	41	67
	30 - 39	10	21	34
	20 - 29	11	11	18

Table F.6.—Machine trades summary, all sites: retest Form B2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Sites	70 - 80	1	72	100
A - G	60 - 69	9	71	99
	50 - 59	17	62	86
	40 - 49	13	45	63
	30 - 39	13	32	44
	20 - 29	17	19	26
	10 - 19	2	2	3

Table F.7.--Vocational/occupational assessment test results by site: mechanical drafting, test Form A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Site H	30 - 40	9	39	100
	20 - 29	27	30	77
	10 - 19	3	3	8
Site I+	50 - 60	10	21	100
	40 - 49	8	11	52
	30 - 39	3	3	14
Site J	50 - 60	2	18	100
	40 - 49	3	16	89
	30 - 39	11	13	72
	20 - 29	2	2	11
Site K	70 - 80	3	19	100
	60 - 69	6	16	84
	50 - 59	9	10	53
	40 - 49	0	1	5
	30 - 39	1	1	5

+ Adjusted scores

Table F.8.--Mechanical drafting summary: all sites, test Form A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Sites H - K	70 - 80	3	97	100
	60 - 69	6	94	97
	50 - 59	21	88	91
	40 - 49	11	67	69
	30 - 39	24	56	58
	20 - 29	29	32	33
	10 - 19	3	3	3

Table F.9.—Vocational/occupational assessment test results by site: mechanical drafting, retest Form B2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Site H	30 - 40	5	21	100
	20 - 29	14	16	76
	10 - 19	2	2	10
Site I+	40 - 50	5	10	100
	30 - 39	4	5	50
	20 - 29	0	1	10
	10 - 19	1	1	10
Site J	50 - 60	1	11	100
	40 - 49	5	10	91
	30 - 39	4	5	45
	20 - 29	1	1	9
Site K	60 - 70	3	12	100
	50 - 59	4	9	75
	40 - 49	3	5	42
	30 - 39	1	2	16
	20 - 29	1	1	8

+ Adjusted scores

Table F.10.--Mechanical drafting summary: all sites, test Form A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Sites H-K	60 - 70	3	54	100
	50 - 59	8	51	94
	40 - 49	10	43	80
	30 - 39	14	33	61
	20 - 29	16	19	35
	10 - 19	3	3	6

Table F.11—Vocational/occupational assessment test results by site: mechanical drafting, retest Form A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Site H	30 - 40	3	18	100
	20 - 29	13	15	83
	10 - 19	2	2	11
Site I+	50 - 60	7	11	100
	40 - 49	3	4	36
	30 - 39	1	1	9
Site J	50 - 60	2	7	100
	40 - 49	3	5	71
	30 - 39	1	2	28
	20 - 29	1	1	14
Site K	60 - 70	6	7	100
	50 - 59	0	1	14
	40 - 49	1	1	14

+Adjusted scores

Table F.12.--Mechanical drafting summary, all sites: retest From A2.

	Scores	Frequency	Cumulative Frequency	Cumulative Percent
Sites H - K	60 - 70	7	43	100
	50 - 59	8	36	84
	40 - 49	7	28	65
	30 - 39	5	21	49
	20 - 29	14	16	37
	10 - 19	2	2	5

Table F.13.—Machine trades mean score summary, all sites: test Form A2, retest Form A2.

Site	Avg. Test Score	Avg. Retest Score	Avg. Test-Retest Difference
A	42.2	43.6	1.4
B	33.0	35.8	2.8
C	43.0	43.8	0.8
D	50.4	47.3	3.1
E	47.1	50.1	3.0
F	58.7	60.1	1.4
G	48.0	42.0	6.0

Table F.14—Machine trades mean score summary, all sites: test Form A2, retest Form B2.

Site	Avg. Test Score	Avg. Retest Score	Avg. Test-Retest Difference
A	36.0	42.2	6.2
B	45.6	29.5	16.1
C	43.3	43.8	0.5
D	49.1	51.0	1.9
E	46.7	39.2	7.5
F	57.2	57.2	0.0
G	50.0	46.5	3.5

Table F.15.--Mechanical drafting mean score summary, all sites: test Form A2, retest Form A2.

Site	Avg. Test Score	Avg. Retest Score	Avg. Test-Retest Difference
H	23.0	25.5	2.5
I	50.0	49.4	0.6
J	44.1	42.0	2.1
K	64.0	63.7	0.3

Table F.16.--Mechanical drafting mean score summary, all sites: test Form A2, retest Form B2.

Site	Avg. Test Score	Avg. Retest Score	Avg. Test-Retest Difference
H	27.3	25.9	1.4
I	45.8	38.6	7.2
J	33.9	40.2	6.3
K	57.5	50.5	7.0

Table F.17.—Vocational/occupational assessment test/retest subscore achievement rankings, by site and duty area: machine trades, test Form A2/retest Form A2.

Site	Duty Area	T/Rt Pass/Pass Frequency	T/Rt Pass/Fail Frequency	T/Rt Fail/Pass Frequency	T/Rt Fail/Fail Frequency
Site A	D	4	1	0	0
	E	0	3	0	2
	F	1	0	0	4
	G	2	1	2	0
	H	0	0	0	5
	I	0	0	0	5
	J	0	0	2	3
	S	1	0	2	2
Site B	C	0	3	2	11
	D	0	4	4	8
	E	0	4	2	10
	F	1	1	3	11
	H	1	0	2	13
	I	0	0	4	12
Site C	C	3	1	1	8
	D	2	5	1	5
	E	4	1	2	6
	F	2	1	0	10
	G	6	0	2	5
	H	0	1	1	11
	I	5	3	1	4
	J	8	2	2	1
Site D	D	7	2	1	0
	E	3	1	2	4
	F	3	3	4	0
	G	2	2	1	5
	H	3	0	4	3
	I	1	5	1	3
	J	3	3	3	1
Site E	C	0	2	2	3
	D	2	1	2	2
	E	3	2	0	2
	F	2	0	0	5
	G	2	1	3	1
	H	3	0	0	4
	I	1	3	1	2
	J	3	1	1	2

Table F.17.—continued.

Duty Area		T/Rt Pass/Pass Frequency	T/Rt Pass/Fail Frequency	T/Rt Fail/Pass Frequency	T/Rt Fail/Fail Frequency
	E	6	1	1	1
Site	F	9	0	0	0
F	G	6	1	2	0
	H	2	0	1	6
	I	3	2	1	3
	J	3	1	2	3
	C	0	0	1	0
	D	1	0	0	0
Site	E	0	0	0	1
G	F	0	0	0	1
	G	1	0	0	0
	H	1	0	0	0
	I	0	0	0	1
	J	0	0	1	0

Table F.18.--Vocational/occupational assessment test/retest subscore achievement rankings summary, by site: machine trades, test Form A2/retest Form A2.

Site	Duties Tested	Test/Retest Pass - Pass Frequrncy (%)	Test/Retest Pass - Fail Frequency (%)	Test/Retest Fail - Pass Frequency (%)	Test/Retest Fail - Fail Frequency (%)
A n=5	D-E-F-G-H-I-J-S	8 (20%)	5 (12%)	6 (15%)	21 (53%)
B n=16	C-D-E-F-H-I	2 (2%)	12 (12%)	17 (18%)	65 (68%)
C n=13	C-D-E-F-G-H-I-J	30 (29)	14 (13%)	10 (10%)	50 (48%)
D n=10	D-E-F-G-H-I-J	22 (31%)	16 (23%)	16 (23%)	16 (23%)
E n=7	C-D-E-F-G-H-I-J	16 (28%)	10 (18%)	9 (16%)	21 (38%)
F n=9	E-F-G-H-I-J	29 (54%)	5 (9%)	7 (13%)	13 (24%)
G n=1	C-D-E-F-G-H-I-J	3 (38%)	0 (0%)	2 (24%)	3 (38%)

Table F.19—Vocational/occupational assessment test/retest subscore achievement rankings, by site and duty area: machine trades, test Form A2/retest Form B2.

Site	Duty Area	T/Rt Pass/Pass Frequency	T/Rt Pass/Fail Frequency	T/Rt Fail/Pass Frequency	T/Rt Fail/Fail Frequency
	D	1	1	2	1
	E	1	2	1	1
Site	F	0	0	0	5
A	G	0	1	3	1
	H	1	0	2	2
	I	1	0	0	4
	J	2	1	0	2
	S	3	0	2	0
	C	2	6	2	11
	D	0	9	2	10
Site	E	0	5	4	12
B	F	1	3	2	15
	H	1	6	1	13
	I	1	5	1	14
	C	4	0	7	1
	D	3	3	1	5
Site	E	3	1	3	5
C	F	2	1	3	6
	G	5	1	2	4
	H	1	1	4	6
	I	1	6	1	4
	J	4	4	1	3
	D	7	1	3	0
	E	2	1	7	1
Site	F	7	3	0	1
D	G	3	1	6	1
	H	3	3	0	5
	I	2	0	2	7
	J	2	7	1	1
	C	4	2	2	0
	D	3	3	0	2
Site	E	3	0	1	4
E	F	2	0	2	4
	G	3	0	1	4
	H	2	2	2	2
	I	3	0	1	4
	J	2	3	0	3

Table F.19.—continued.

Duty Areas		T/Rt Pass/Pass	T/Rt Pass/Fail	T/Rt Fail/Pass	T/Rt Fail/Fail
	E	11	2	0	0
	F	6	5	0	2
Site	G	7	2	3	1
F	H	3	3	5	2
	I	2	1	6	4
	J	7	0	4	2
	C	1	0	1	0
	D	0	1	0	1
Site	E	0	1	0	1
G	F	0	0	1	1
	G	0	0	1	1
	H	0	0	1	1
	I	0	0	1	1
	J	0	0	0	2

Table F.20.--Vocational/occupational assessment test/retest subscore achievement rankings summary, by site: machine trades, test Form A2/retest Form B2.

Site	Duties Tested	Test/Retest Pass - Pass Frequency (%)	Test/Retest Pass - Fail Frequency (%)	Test/Retest Fail - Pass Frequency (%)	Test/Retest Fail - Fail Frequency (%)
Site A	D-E-F G-H-I J-S	9 (23%)	5 (12%)	10 (25%)	16 (40%)
Site B	C-D-E F-H-I	5 (4%)	34 (27%)	12 (10%)	75 (59%)
Site C	C-D-E G-H-I I-J	23 (24%)	17 (18%)	22 (23%)	34 (35%)
Site D	D-E-F G-H-I J	26 (33%)	16 (21%)	19 (25%)	16 (21%)
Site E	C-D-E F-G-H I-J	22 (24%)	10 (16%)	9 (14%)	23 (36%)
Site F	E-F-G H-I-J	36 (46%)	13 (17%)	18 (23%)	11 (14%)
Site G	C-D-E F-G-H I-J	1 (6%)	2 (13%)	5 (31%)	8 (50%)

Table F.21.--Vocational/occupational assessment test/retest subscore achievement rankings, by site and duty area: mechanical drafting, test Form A2/retest Form A2.

	Duty Area	T/Rt Pass/Pass Frequency	T/Rt Pass/Fail Frequency	T/Rt Fail/Pass Frequency	T/Rt Fail/Fail Frequency
Site H	F	0	0	0	18
	G	0	0	0	18
	J	0	0	0	18
	L	0	0	2	16
	M	0	1	0	17
	N	0	2	2	14
Site I	F	10	0	0	1
	G	9	2	0	0
	J	4	1	4	2
	L	2	4	3	2
	M	6	2	0	3
	N	6	0	2	3
Site J	F	0	0	0	7
	G	2	3	0	2
	J	1	1	1	4
	L	2	0	1	4
	M	3	0	2	2
	N	1	3	0	3
Site K	F	7	0	0	0
	G	6	0	1	0
	H	4	0	2	1
	I	6	0	0	1
	J	7	0	0	0
	L	7	0	0	0
	M	5	1	0	1
	N	0	1	2	4

Table F.22.--Vocational/occupational assessment test/retest subscore achievement rankings, summary by site: mechanical drafting, test Form A2/retest Form A2.

	Duties Tested	Test/Retest Pass - Pass Frequency (%)	Test/Retest Pass - Fail Frequency (%)	Test/Retest Fail - Pass Frequency (%)	Test/Retest Fail - Fail Frequency (%)
Site H	F-G-J L-M-N	0 (0%)	3 (2%)	4 (4%)	101 (94%)
Site I	F-G-J L-M-N	37 (56%)	9 (14%)	9 (14%)	11 (16%)
Site J	F-G-J L-M-N	9 (21%)	7 (17%)	4 (10%)	22 (52%)
Site K	F-G-H I-J-L M-N	42 (75%)	2 (4%)	5 (9%)	7 (12%)

Table F.23.--Vocational/occupational assessment test/retest subscore achievement rankings, by site and duty area: mechanical drafting, test Form A2/retest Form B2.

	Duty Area	T/Rt Pass/Pass Frequency	T/Rt Pass/Fail Frequency	T/Rt Fail/PAss Frequency	T/Rt Fail/Fail Frequency
	F	0	0	1	20
	G	0	0	0	21
Site	J	0	0	0	21
H	L	0	0	1	20
	M	0	5	0	16
	N	0	8	0	13
	F	2	0	6	2
	G	3	5	0	2
Site	J	1	2	0	7
I	L	2	5	0	3
	M	2	3	1	4
	N	2	2	3	3
	F	0	1	3	7
	G	0	0	4	7
Site	J	0	0	3	8
J	L	2	0	3	6
	M	0	1	1	9
	N	1	1	1	8
	F	9	0	3	0
	G	5	1	5	1
Site	H	2	2	7	1
K	I	8	2	2	0
	J	7	3	2	0
	L	9	2	0	1
	M	3	4	3	2
	N	3	0	3	6

Table F.24.—Vocational/occupational assessment test/retest subscore achievement rankings, summary by site: mechanical drafting, test Form A2/retest Form A2.

	All Duties Tested	Test/Retest Pass - Pass Frequency (%)	Test/Retest Pass - Fail Frequency (%)	Test/Retest Fail - Pass Frequency (%)	Test/Retest Fail - Fail Frequency (%)
Site H	F-G-J L-M-N	0 (0%)	3 (2%)	4 (4%)	101 (94%)
Site I	F-G-J L-M-N	37 (56%)	9 (14%)	9 (14%)	11 (16%)
Site J	F-G-J L-M-N	9 (21%)	7 (17%)	4 (10%)	22 (52%)
Site K	F-G-H I-J-L M-N	42 (75%)	2 (4%)	5 (9%)	7 (12%)

Table F.25.--Vocational/occupational assessment test-retest subscore agreement and correlational data: machine trades, test Form A2/retest Form A2.

	Duty Areas	Percent Agreement Pass-Pass/Fail-Fail	Percent Disagreement Pass-Fail/Fail-Pass	Tetrachoric Correlation
Site A	D	80	20	----
	E	40	60	----
	F	100	0	1.0
	G	40	60	----
	H	100	0	1.0
	I	100	0	1.0
	J	60	40	----
	S	60	40	----
Avg.D-S		73	27	----
Site B	C	69	31	----
	D	50	50	----
	E	63	37	----
	F	75	25	.410
	H	88	12	----
	I	75	25	----
Avg.C-I		70	30	----
Site C	C	85	15	----
	D	54	46	----
	E	77	23	----
	F	92	8	----
	G	85	15	----
	H	85	15	----
	I	69	31	----
	J	69	31	----
Avg.C-J		77	23	----
Site D	D	70	30	----
	E	70	30	.607
	F	30	70	----
	G	70	30	.549
	H	60	40	----
	I	40	60	.181
	J	40	60	.397
Avg.D-J		54	46	----

Table F.25.--continued.

	Duty Area	Percent Agreement Pass-Pass/Fail-Fail	Percent Disagreement Pass-Fail/Fail-Pass	Tetrachoric Correlation
	C	43	57	----
	D	57	43	----
Site	E	72	28	----
E	F	100	0	1.0
	G	43	57	.151
	H	100	0	1.0
	I	43	57	.151
	J	72	28	.611
Avg.C-J		66	33	----
	E	78	22	.574
Site	F	100	0	1.0
F	G	67	33	----
	H	89	11	----
	I	67	33	.529
	J	67	33	.529
Avg. E-J		78	22	----

Table F.26.--Vocational/occupational assessment test-retest subscore agreement and correlational data: mechanical drafting, test Form A2/retest Form A2.

	Duty Areas	Percent Agreement Pass-Pass/Fail-Fail	Percent Disagreement Pass-Fail/Fail-Pass	Tetrachoric Correlation
	F	100	0	1.0
Site	G	100	0	1.0
H	J	100	0	1.0
	L	89	11	----
	M	94	6	----
	N	78	22	----
Avg.	F-N	94	6	----
	F	90	10	----
	G	82	18	----
Site	J	55	45	----
I	L	36	64	----
	M	82	18	----
	N	82	18	----
Avg.	F-N	71	29	----
	F	100	0	1.0
	G	57	43	----
Site	J	71	29	----
J	L	86	14	----
	M	71	29	----
	N	57	43	----
Avg.	F-N	71	29	----
	F	100	0	1.0
	G	86	14	.607
Site	H	71	29	----
K	I	100	0	1.0
	J	100	0	1.0
	L	100	0	1.0
	M	86	14	----
	N	57	43	----
Avg.	F-N	88	12	----

Table F.27.--Vocational/occupational assessment test-retest subscore agreement and correlational data: machine trades, test Form A2/retest Form B2.

	Duty Areas	Percent Agreement Pass-Pass/Fail-Fail	Percent Disagreement Pass-Fail/Fail-Pass	Tetrachoric Correlation
	D	40	60	.262
Site	E	40	60	.262
A	F	100	0	1.0
	G	20	80	----
	H	60	40	----
	I	100	0	1.0
	J	80	20	----
	S	60	40	----
Avg. D-S		63	37	----
	C	62	38	.137
Site	D	48	52	----
B	E	57	43	----
	F	76	24	.291
	H	67	33	----
	I	71	29	----
Avg. C-I		64	36	----
	C	42	58	----
Site	D	67	33	.549
C	E	67	33	.549
	F	67	33	.475
	G	75	25	.723
	H	58	42	.140
	I	42	58	.140
	J	58	42	.391
Avg. C-J		60	40	----
	D	64	36	----
Site	E	27	73	.417
D	F	73	27	----
	G	36	64	.240
	H	72	28	----
	I	82	18	----
	J	27	73	----
Avg. D-J		54	46	----

Table F.27--continued.

	Duty Areas	Percent Agreement Pass-Pass/Fail-Fail	Percent Disagreement Pass-Fail/Fail-Pass	Tetrachoric Correlation
	C	50	50	----
Site	D	63	37	----
E	E	88	12	----
	F	75	25	----
	G	88	12	----
	H	50	50	----
	I	88	12	----
	J	63	37	----
Avg.	C-J	71	29	----
	E	85	15	----
Site	F	62	38	----
F	G	62	38	.054
	H	39	62	.342
	I	46	54	.102
	J	69	31	----
Avg.	E-J	61	39	----

Table F.28.—Vocational/occupational assessment test/retest subscore agreement and correlational data: mechanical drafting, test Form A2/retest Form B2.

	Duty Areas	Percent Agreement Pass-Pass/Fail-Fail	Percent Disagreement Pass-Fail/Fail-Pass	Tetrachoric Correlations
	F	95	5	----
Site	G	100	0	1.0
H	J	100	0	1.0
	L	95	5	----
	M	76	24	----
	N	62	38	----
Avg.	F-N	88	12	----
	F	40	60	----
Site	G	50	50	----
I	J	80	20	----
	L	50	50	----
	M	60	40	----
	N	50	50	----
Avg.	F-N	55	45	----
	F	64	36	----
Site	G	64	36	----
J	J	73	27	----
	L	73	27	----
	M	82	18	----
	N	82	18	----
Avg.	F-N	73	27	----
	F	75	25	----
Site	G	50	50	----
K	H	25	75	.627
	I	67	33	----
	J	58	42	----
	L	83	17	----
	M	42	58	----
	N	75	25	----
Avg.	F-N	58	42	----

Table F.29.—Vocational/occupational assessment test subscore aveage in relation to instructors' subscore rankings, by site and duty Area: machine trades.

	Duty Area	INSTRUCTOR RANKING Pass/Pass Frequency	Pass/Fail Frequency	DUTY SUBSCORE RANKING Fail/Pass Frequency	Fail/Fail Frequency
	D	6	3	0	1
Site	E	2	6	1	1
A	F	0	0	1	9
	G	1	3	2	4
	H	0	3	2	5
	I	0	4	1	5
	J	5	5	0	0
	S	5	0	2	3
	C	2	6	7	22
Site	D	4	7	5	21
B	E	6	28	0	3
	F	7	29	0	1
	H	8	26	0	3
	I	2	7	8	20
	C	10	12	0	3
Site	D	8	4	6	7
C	E	8	7	1	9
	F	6	18	0	1
	G	1	1	13	10
	H	2	8	0	15
	I	0	2	8	15
	J	16	7	1	1
	D	15	3	1	2
Site	E	14	7	0	0
D	F	14	7	0	0
	G	6	5	4	6
	H	8	9	0	4
	I	4	11	0	6
	J	15	6	0	0
	C	1	0	4	10
Site	D	7	4	1	3
E	E	8	4	1	2
	F	5	6	0	4
	G	2	2	5	6
	H	4	3	3	5
	I	5	1	2	7
	J	8	5	0	2

Table F.29—continued.

	Duty Area	INSTRUCTOR RANKING—DUTY SUBSCORE RANKING Pass/Pass Frequency	Pass/Fail Frequency	Fail/Pass Frequency	Fail/Fail Frequency
	E	19	3	0	0
Site	F	20	2	0	0
F	G	17	3	0	2
	H	3	4	5	10
	I	12	8	0	2
	J	16	5	0	1
	C	1	2	0	0
Site	D	1	2	0	0
G	E	0	3	0	0
	F	1	2	0	0
	G	1	2	0	0
	H	1	2	0	0
	I	0	3	0	0
	J	1	2	0	0

Table F.30.--Vocational/occupational assessment test subscore aveage in relation to instructors' subscore rankings, by site and duty area: mechanical drafting.

	Duty Area	INSTRUCTOR RANKING------DUTY SUBSCORE RANKING Pass/Pass Frequency	Pass/Fail Frequency	Fail/Pass Frequency	Fail/Fail Frequency
Site H	F	0	0	0	39
	G	0	0	0	39
	J	0	0	0	39
	L	0	0	2	37
	M	0	0	0	39
	N	0	0	0	39
Site I	F	0	0	14	7
	G	0	0	15	6
	J	5	9	2	5
	L	11	9	0	1
	M	9	11	0	1
	N	8	13	0	0
Site J	F	0	0	1	17
	G	0	0	5	13
	J	1	17	0	0
	L	5	5	0	8
	M	5	13	0	0
	N	3	12	1	2
Site K	F	17	1	0	1
	G	4	1	12	2
	H	6	8	2	3
	I	7	0	9	3
	J	14	1	3	1
	L	8	0	9	2
	M	15	4	0	0
	N	3	16	0	0

Table F.31.--Vocational/occupational assessment test subscore average in relation to instructors' rankings, summary by site: machine trades.

	Duties Tested	INSTRUCTOR RANKING--DUTY SUBSCORE RANKING Pass/Pass Frequency(%)	Pass/Fail Frequency(%)	Fail/Pass Frequency(%)	Fail/Fail Frequency(%)
Site A	D-E-F H-I-J J-S	19 (13%)	24 (30%)	9 (11%)	28 (46%)
Site B	C-D-E F-H-I	29 (13%)	103 (46%)	20 (9%)	70 (32%)
Site C	C-D-E F-G-H I-J	51 (25%)	59 (30%)	29 (15%)	61 (30%)
Site D	D-E-F G-H-I J	76 (52%)	48 (33%)	5 (3%)	18 (12%)
Site E	C-D-E F-G-H I-J	40 (33%)	25 (21%)	16 (13%)	39 (33%)
Site F	E-F-G H-I-J	87 (66%)	25 (19%)	5 (4%)	15 (11%)
Site G	C-D-E F-G-H I-J	6 (25%)	18 (75%)	0 (0%)	0 (0%)

Table F.32.--Vocational/occupational assessment test subscore averge in relation to instructors' rankings, summary by site: mechanical drafting.

	Duties Tested	INSTRUCTOR RANKING---Pass-Pass Frequency(%)	Pass-Fail Frequency(%)	Fail-Pass Frequency(%)	DUTY SUBSCORE RANKING---Fail-Fail Frequency(%)
Site H	F-G-J L-M-N	0 (0%)	0 (0%)	2 (1%)	232 (99%)
Site I	F-G-J L-M-N	33 (26%)	42 (33%)	31 (25%)	20 (16%)
Site J	F-G-J L-M-N	14 (13%)	47 (44%)	7 (6%)	40 (37%)
Site K	F-G-H I-J-L M-N	74 (49%)	31 (20%)	35 (23%)	12 (8%)

Table F.33.--Vocational/occupational assessment test/retest, pass/fail achievement, for machine trades, all sites: test Form A2/retest Form A2.

		TEST A2 Pass	TEST A2 Fail
RETEST A2	Pass	110 (26%)	67 (16%)
	Fail	62 (14%)	189 (44%)

Table F.34.--Vocational/occupational assessment test/retest pass/fail achievement for machine trades, all sites: test Form A2/retest Form B2.

		TEST A2 Pass	TEST A2 Fail
RETEST B2	Pass	122 (25%)	95 (19%)
	Fail	97 (20%)	183 (36%)

Table F.35.--Vocational/occupational assessment test/retest pass/fail achievement for mechanical drafting, all sites: test Form A2/retest Form A2.

	TEST A2 Pass	TEST A2 Fail
RETEST A2 Pass	88 (32%)	22 (8%)
RETEST A2 Fail	21 (8%)	141 (52%)

Table F.36.--Vocational/occupational assessment test/retest pass/fail achievement for mechanical drafting, all sites: test Form A2/retest Form B2.

	TEST A2 Pass	TEST A2 Fail
RETEST B2 Pass	61 (18%)	46 (13%)
RETEST B2 Fail	53 (15%)	188 (54%)

Tables F.37-F.42.--Machine trades student percentile rank (achievement) in relation to instructor's ranking of students: machine trades, all sites, in frequency and (percent).

Table F.37--Site A

Instructor Ranking

Actual Ranking	1	2	3	4	5
1			1 (10%)		
2				3 (30%)	
3				3 (30%)	
4				2 (20%)	
5					1 (10%)

Note: Percentage agreements = 30
Student achievement above rating = 0 percent
Student achievement below rating = 70 percent

Table F.38--Site B

Instructor Rating

Actual Rating	1	2	3	4	5
1		2 (5.4%)	2 (5.4%)		1 (2.7%)
2	2 (5.4%)	2 (5.4%)	4 (10.8%)	1 (2.7%)	1 (2.7%)
3		2 (5.4%)		4 (10.8%)	
4		1 (2.7%)	3 (8.1%)	3 (8.1%)	1 (2.7%)
5			3 (8.1%)	3 (8.1%)	2 (5.4%)

Note: Percentage agreements = 18.9
Student achievement above rating = 29.7 percent
Student achievement below rating = 51.4 percent

Table F.39.--Site C

Instructor Rating

Actual Rating	1	2	3	4	5
1		2 (8%)	3 (12%)		
2	1 (4%)	1 (4%)	1 (4%)		
3			5 (20%)	1 (4%)	1 (4%)
4			3 (12%)	3 (12%)	
5			1 (4%)	1 (4%)	2 (8%)

Note: Percentage agreements = 44
Student achievement above rating = 24 percent
Student achievement below rating = 32 percent

Table F.40.--Site D

Instructor Rating

Actual Rating	1	2	3	4	5
1			1 (4.8%)		1 (4.8%)
2				5 (23.8%)	2 (9.5%)
3				2 (9.5%)	2 (9.5%)
4					4 (19%)
5					4 (19%)

Note: Percentage agreements = 19
Student achievement above rating = 0
Student achievement below rating = 81

Table F.41.--Site E

Instructor Rating

Actual Rating	1	2	3	4	5
1		3 (20%)			
2		3 (20%)	1 (6.7%)		
3			1 (6.7%)		
4			1 (6.7%)	3 (20%)	
5				2 (13.3%)	1 (6.7%)

Note: Percentage agreements = 53.4
Student achievement above rating = 20 percent
Student achievement below rating = 26.6 percent

Table F.42.--Site F

Instructor Rating

Actual Rating	1	2	3	4	5
1	1 (4.5%)				
2			5 (22.7%)		1 (4.5%)
3			2 (9.1%)	3 (13.6%)	1 (4.5%)
4				1 (4.5%)	4 (18.2%)
5				2 (9.1%)	2 (9.1%)

Note: Percentage agreements = 27.2
Student achievement above rating = 9.1 percent
Student achievement below rating = 63.7 percent

Table F.43.--Student percentile rank achievement in relation to instructors' rankings of students: machine trades, Sites A-F combined, in frequency and (percent).

Sites A - F

Instructor Rating

Actual Rating	1	2	3	4	5
1	1 (.5%)	7 (5%)	7 (5%)		2 (1%)
2	3 (2%)	6 (4.5%)	11 (8%)	9 (7%)	5 (4%)
3		2 (1%)	8 (6%)	13 (10%)	5 (4%)
4		1 (.5%)	7 (5%)	12 (9%)	9 (7%)
5			4 (3%)	8 (6%)	13 (10%)

Note: Percentage agreements = 30
Student achievement above rating = 19 percent
Student achievement below rating = 51 percent

Tables F.44-F.47.--Student percentile rank in relation to instructors' rankings of students: mechanical drafting, all sites, in frequency and (percent).

Table F.44--Site H

Instructor Rating

Actual Rating	1	2	3	4	5
1	4 (10.3%)	1 (2.6%)			
2	3 (7.7%)	2 (5.1%)	2 (5.1%)	2 (5.1%)	
3	2 (5.1%)	7 (17.9%)	5 (12.8%)		
4	2 (5.1%)	1 (2.6%)	3 (7.7%)	2 (5.1%)	
5			2 (5.1%)	1 (2.6%)	

Note: Percentage agreements = 33.3
Student achievement above rating = 53.8 percent
Student achievement below rating = 12.9 percent

Table F.45--Site I

Instructor Rating

Actual Rating	1	2	3	4	5
1			1 (4.8%)	2 (9.5%)	
2				4 (19%)	1 (4.8%)
3				4 (19%)	2 (9.5%)
4			1 (4.8%)	2 (9.5%)	2 (9.5%)
5				1 (4.8%)	1 (4.8%)

Note: Percentage agreements = 14.3
Student achievement above rating = 9.6 percent
Student achievement below rating = 76.1 percent

Table F.46.--Site J

Instructor Rating

Actual Rating	1	2	3	4	5
1		1 (5.6%)			
2		1 (5.6%)	3 (16.7%)	3 (16.7%)	
3			1 (5.6%)	1 (5.6%)	1 (5.6%)
4				4 (22.2%)	
5					3 (16.7%)

Note: Percentage agreements = 50.1
Student achievement above rating = 0 percent
Student achievement below rating = 49.9 percent

Table F.47.—Site K

Instructor Rating

Actual Rating	1	2	3	4	5
1			1 (5.3%)	1 (5.3%)	1 (5.3%)
2				1 (5.3%)	2 (10.5%)
3				2 (10.5%)	2 (10.5%)
4			1 (5.3%)	2 (10.5%)	2 (10.5%)
5				1 (5.3%)	3 (15.8%)

Note: Percentage agreements = 26.3
Student achievement above rating = 10.6 percent
Student achievement below rating = 63.1 percent

Table F.48.—Student percentile rank in relation to instructors' rankings of student rank: mechanical drafting, sites H-K combined, in frequency and (percent).

Sites H - K

Instructor Rating

Actual Rating	1	2	3	4	5
1	4 (4%)	2 (2%)	2 (2%)	3 (3%)	1 (1%)
2	3 (3%)	3 (3%)	5 (5%)	10 (10%)	3 (3%)
3	2 (2%)	7 (7%)	6 (6%)	7 (7%)	5 (5%)
4	2 (2%)	1 (1%)	5 (5%)	10 (10%)	4 (4%)
5			2 (2%)	3 (3%)	7 (7%)

Note: Percentage agreements = 30.9
Student achievement above rating = 25 percent
Student achievement below rating = 44.1 percent

REFERENCES

REFERENCES

American Vocational Association. Fact Sheet. Arlington, Va.: American Vocational Association, March 1984.

Apple, Michael. "Curriculum in the Year 2000: Tensions and Possibilities." Educational Leadership (January 1983).

Barr, Arvil S. "Research Methods." In Encyclopedia of Educational Research, pp. 1160-66. Edited by Chester W. Harris. New York: Macmillan, 1960.

Bloom, Benjamin S.; Hastings, Thomas J.; and Madaus, George F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill Book Co., 1971.

Borg, W. R., and Gall, M. D. Educational Research: An Introduction. New York: Longman, 1983.

Bottoms, Gene. "Moving Toward Excellence." American Vocational Education Journal (November/December 1984).

Brookover, Wilbur; Beamer, Laurence; Efthim, Helen; Hathaway, Douglas; Lezotte, Lawrence; Miller, Stephen; Passalacque, Joseph; and Tornatzky, Louis. Creating Effective Schools: An Inservice Program for Enhancing School Learning Climate and Achievement. Holmes Beach, Fla.: Learning Publications, 1982.

Carver, R. P. "Special Problems in Measuring Change With Psychometric Devices." In Evaluation Research: Strategies and Methods. Washington, D.C.: American Institute for Research, 1970.

Cawelti, Gordon L. Effective Schools Movement and Evaluation. Association for Supervision and Curriculum Development.

Chamberlin, Dean; Chamberlin, Enid; Drought, N. D., and Scott, W. E. Did They Succeed in College? New York: Harper and Brothers, 1942.

Cronbach, Lee J. "Course Improvement Through Evaluation." In Curriculum Evaluation. Edited by David A. Payne. Lexington, Mass.: D. C. Heath, 1974.

Dressel, Paul L. Evaluation as Instruction. Proceedings of the 1953 Invitational Conference on Testing Programs. Princeton, N.J.: Educational Testing Service, 1954.

Durant, William. The Life of Greece. New York: Simon & Schuster, 1939.

Eaddy, Kenneth. Director of V-TECS. Telephone interview, February 25, 1986.

Ebel, Robert L. Essentials of Educational Measurements. 3d ed. Englewood Cliffs, N.J.: Prentice-Hall, 1979.

Edmonds, Ronald R. "Effective-Schools Correlates." Education Week, January 15, 1986.

Evans, Rupert N., and Herr, Edwin L. Foundations of Vocational Education. Columbus, Ohio: Charles E. Merrill, 1978.

Faggen, Jane. Decision Reliability and Classification Validity for Decision Oriented Criterion-Referenced Tests. Bethesda, Md.: ERIC Document Reproduction Service, ED 159 221, 1979.

Flanagan, J. C.; Davis, F. B.; Dailey, J. T.; Shaycroft, M. F.; Orr, D. B.; Goldberg, I.; and Neyman, C. A., Jr. Project TALENT. The Identification, Development, and Utilization of Human Talents: The American High School Student. Pittsburgh: University of Pittsburgh Press, 1964.

Finch, C. R., and Crunkilton, J. R. Curriculum Development in Vocational and Technical Education. Boston: Allyn and Bacon, 1979.

Fox, Robert L. The Research Process in Education. New York: Holt, Rinehart, and Winston, 1969.

Glaser, Robert, and Nitko, Anthony J. "Measurement in Learning and Instruction." In Educational Measurement. 2d ed. Edited by Robert L. Thorndike. Washington, D.C.: American Council on Education, 1971.

Gleason, James. Director, IDECC. Telephone interview, March 1986.

Harnisch, Delwyn L. Computer Application Issues in Certification and Licensure Testing. Bethesda, Md.: ERIC Document Reproduction Service, ED 261 079, 1986.

Harris, Chester W. Techniques for Analyzing Test Response Data. Bethesda, Md.: ERIC Document Reproduction Service, ED 110 496, 1975.

Hillway, Tyrus. Introduction to Research. Boston: Houghton-Mifflin, 1956.

Humbleton, Ronald K., and Novick, Melvin R. "Toward an Integration of Theory and Methods for Criterion-Referenced Tests." Journal of Educational Measurement 10 (Fall 1973).

Independent School. A Nation at Risk: The Imperative for Educational Reform. An Open Letter to the American People. Independent School, October 1983.

Kimpston, Richard D., and Anderson, Douglas H. "A Study to Analyze Curriculum Decision Making in School Districts." Educational Leadership 40 (November 1982): 63-66.

Klein, Stephen P., and Kosecoff, Jacqueline. "Issues and Procedures in the Development of Criterion-Referenced Tests." In Readings in Measurement and Evaluation in Education and Psychology. Edited by William A. Mehrens. New York: Holt, Rinehart, and Winston, 1976.

Koch, William R., and Reckase, Mark D. Problems in Application of Latent Trait Models to Tailored Testing. Bethesda, Md.: ERIC Document Reproduction Service, ED 177 196, 1980.

Kromhout, Ora. Center for Instructional Development and Services, Tallahassee, Florida. Telephone interview, February 15, 1986.

Lathrop, Robert. Center for Instructional Development and Services, Tallahassee, Florida. Telephone interview.

Lee, Connie W. Third Progress and Information Report of the Vocational-Technical Education Consortium of States. Atlanta: Southern Association of Colleges and Schools, July 1976.

Lewis, Morgan V., and Martin, Scott. Measures of Occupationally Specific and Nonoccupationally Specific Knowledge and Skills: A Compendium. Columbus: National Center for Research in Vocational Education, The Ohio State University, 1986.

Lezotte, Lawrence. "On the Success of the Effective Schools Movement by Lynn Olson." Education Week, January 15, 1986.

Martuza, V. R. Applying Norm-Referenced and Criterion-Referenced Measures in Education. Boston: Allyn and Bacon, 1977.

McCage, Ronald. [Vocational-Technical Education Consortium of States] Newsletter to Member States. July 1986.

Mehrens, William A., and Lehmann, Irvin J. Measurement and Evaluation in Education and Psychology. New York: Holt, Rinehart and Winston, 1984.

Meranto, Phillip. "The Politics of Federal Aid to Education in 1965." In A History of Congressional Acts and Court Decisions. Ann Arbor: University of Michigan Press, 1974.

Michigan Department of Education. Competency-Based Education In-Service Modules. Lansing: Department of Education, 1974.

________. Vocational-Technical Education Service. Memo from the Superintendent of Schools, Phillip Runkel. February 11, 1986.

Miller, P. L. "Change Agent Strategies: A Study of the Michigan-Ohio Regional Educational Laboratory." Dissertation Abstracts International 29 (1968): 1684A. (University Microfilms No. 68-17, 112)

Millman, J. "Criterion-Referenced Measurement." In Evaluation in Education, Current Applications. Edited by W. J. Popham. Berkeley, Calif.: McCutchan, 1974.

Morsh, J. E., and Archer, W. B. Procedural Guide for Conducting Occupational Surveys in the United States Air Force. Lackland Air Force Base, Texas: Personnel Research Laboratory, September 1967.

Naisbitt, John. Megatrends. New York: Warner Books, 1984.

National Center for Educational Statistics. Vocational Education Data Survey. Preliminary statistics for 1980-1981.

National Institute of Education. Survey Results on the Educational Attitudes of the American People. Washington, D.C.: National Institute of Education, 1985.

Oberstein, Diane R. "Item Bank Utilization." Ph.D. dissertation, University of California at Los Angeles, 1981.

Oliver, Dale. Professor of Education, Division of Vocational-Technical Education, Virginia Polytechnic Institute and State University. Telephone interview, February 16, 1986.

Olson, Chris. Projects Coordinator, Vocational Education Curriculum Project, Michigan State University. Personal interview, November 1985.

Olson, Lynn. "Effective Schools." Education Week, January 15, 1986.

O'Reilly, Robert P., and others. Validation of a Multiple-Choice Cloze Test of Literal Comprehension: Summary Report. Bethesda, Md.: ERIC Document Reproduction Service, ED 120 265, 1976.

Parnell, Dale. Hearings before the General Subcommittee on Education of the Committee on Education and Labor, House of Representatives, 93rd Cong., 1st sess. In Elementary and Secondary Education Amendments of 1973.

Pollock, William T., and McDole, Thomas L. Development of a National Item Bank for Tests of Driving Knowledge. Bethesda, Md.: ERIC Document Reproduction Service, ED 128 403, 1977.

Popham, James, and Husek, T. R. "Implications of Criterion-Referenced Measurement." Journal of Educational Measurement 6 (Spring 1969): 1-9.

Pratt, William. "Michigan Council of Vocational Administrators' Response to A Nation at Risk." November 1983.

Runkel, Phillip. "Approval of Secondary Vocational Program Evaluation Model." Lansing: State Board of Education, February 25, 1986.

Sax, Gilbert. Principles of Educational Measurement. Belmont, Calif.: Wadsworth Publishing Co., 1974.

Scriven, Michael. "The Methodology of Evaluation." In Curriculum and Evaluation. Edited by Arno A. Bellack and Herbert M. Kliebard. Berkeley, Calif.: McCutchan Publishing Corp., 1977.

Seney, Mack. Michigan Department of Education. Personal interview, March 7, 1986.

Smith, Nick, U. S. Senator. "Accountability in Education." Senate Bill 195. Lansing, Mich.: Senate Analysis Section, 1986.

Stroud, J. Psychology in Education. New York: McKay, 1946.

Stufflebeam, D. L.; Foley, W. J.; Gephart, W. T.; Guba, E. G.; Hammond, R. L.; Herriman, H. O.; and Provus, M. M. Educational Evaluation and Decision-Making in Education. Itasca, Ill.: Peacock, 1971.

Taylor, Fredrick. Scientific Management. New York: Harper and Row, 1947.

Thorndike, R. L., and Hagen, E. Measurement and Evaluation in Psychology and Education. New York: Wiley, 1969.

Tyler, Ralph. Basic Principles of Curriculum and Instruction. Chicago: University of Chicago Press, 1949.

U.S. Congress. The Vocational Education Act of 1963. Public Law 88-210. Washington, D.C.: Government Printing Office, 1963.

________. Elementary and Secondary Education Act of 1965. Public Law 89-10. Washington, D.C.: Government Printing Office, 1965.

________. The Vocational Education Amendments of 1968. Public Law 95-76. Washington, D.C.: Government Printing Office, 1968.

________. The 1976 Vocational Education Amendments. Public Law 94-482. Washington, D.C.: Government Printing Office, 1976.

________. The Carl Perkins Vocational Education Act. Public Law 98-524. Washington, D.C.: Government Printing Office, 1984.

U.S. Department of Education. National Commission on Excellence in Education. A Nation at Risk. Washington, D.C.: Government Printing Office, 1983.

Vocational-Technical Education Consortium of States. Handbook for Developing Criterion-Referenced Test Items. Atlanta: V-TECS, 1985.

Wentling, Tim L. Evaluating Occupational Education and Training Programs. Urbana, Ill.: Griffon Press, 1980.