DOCUMENT RESUME

ED 030 734 VT 008 714

By-Koenigsberg, Lewis A., Reilly, Robert R.

An Investigation of the Reliability and Validity of Selected Occupational Competency Examinations and Their Use in Evaluating Prospective Trade and Industrial Teachers. Final Report.

New York State Education Dept., Albany. Bureau of Occupational Education Research; State Univ. of New York, Oswego. Div. of Vocational Technical Education.
Report No-BOR-8
Pub Date Jun 68
Note-45p.
EDRS Price MF-$0.25 HC-$2.35
Descriptors-Auto Mechanics (Occupation), Certification, Cosmetology, Item Analysis, Machinists, Occupational Tests, *Performance Tests, *Teacher Qualifications, Test Construction, *Test Reliability, *Test Validity, *Trade and Industrial Teachers
Identifiers-New York State Education Department

A 2-year investigation was initiated to determine the reliability and validity of the New York State competency tests used in selecting candidates for teacher preparation in the trade and industrial programs. The three most widely used tests--auto mechanics, cosmetology, and machine shop--were selected for investigation and further revision. Several statistical approaches were employed to test reliability of the written tests. Some findings were: (1) The mean difficulty of items was within the range considered acceptable. (2) The mean point biserial was lower for all tests than generally recommended. (3) The variability of difficulty levels was extremely high. On performance tests, comparison between judges rating applicants on performance showed a high relationship and intercorrelations among sub-scores were high. Qualified individuals were used to determine whether items reflect the important aspects of knowledge or skill the tests were designed to measure, and the relationship between examination scores and actual experience in the occupation was examined. Recommendations included: (1) pooling test data from all parts of the state, and (2) a plan for systematic review and updating. (JK)

RESEARCH COORDINATING UNIT

FINAL REPORT

BOR-8

AN INVESTIGATION OF THE RELIABILITY AND VALIDITY OF SELECTED OCCUPATIONAL COMPETENCY EXAMINATIONS AND THEIR USE IN EVALUATING PROSPECTIVE TRADE AND INDUSTRIAL TEACHERS

The State University of New York
College of Arts and Sciences
Oswego, New York

In Cooperation With

The University of the State of New York
THE STATE EDUCATION DEPARTMENT
Bureau of Occupational Education Research

June 1968

AN INVESTIGATION OF THE RELIABILITY AND VALIDITY OF SELECTED OCCUPATIONAL COMPETENCY EXAMINATIONS AND THEIR USE IN EVALUATING PROSPECTIVE TRADE AND INDUSTRIAL TEACHERS

by

Lewis A. Koenigsberg

and

Robert R. Reilly

Evaluation Consultants for

The Division of Vocational Technical Education
College of Arts and Sciences
Oswego, New York

June 1968

THE UNIVERSITY OF THE STATE OF NEW YORK

Regents of the University

(with years when terms expire)

1969	Joseph W. McGovern, A.B., LL.B., L.H.D., LL.D., *Chancellor*	New York
1970	Everett J. Penny, B.C.S., D.C.S., *Vice Chancellor*	White Plains
1978	Alexander J. Allan, Jr., LL.D., Litt.D.	Troy
1973	Charles W. Millard, Jr., A.B., LL.D., L.H.D.	Buffalo
1972	Carl H. Pforzheimer, Jr., A.B., M.B.A., D.C.S.	Purchase
1975	Edward M. M. Warburg, B.S., L.H.D.	New York
1977	Joseph T. King, LL.B.	Queens
1974	Joseph C. Indelicato, M.D.	Brooklyn
1976	Mrs. Helen B. Power, A.B., Litt.D., L.H.D.	Rochester
1979	Francis W. McGinley, B.S., LL.B., LL.D.	Glens Falls
1981	George D. Weinstein, LL.B.	Hempstead
1980	Max J. Rubin, LL.B., L.H.D.	New York
1971	Kenneth B. Clark, A.B., M.S., Ph.D., Litt.D.	Hastings on Hudson
1982	Stephen K. Bailey, A.B., B.A., M.A., Ph.D., LL.D.	Syracuse
1983	Harold E. Newcomb, B.A.	Owego

President of the University and Commissioner of Education
James E. Allen, Jr.

Executive Deputy Commissioner of Education
Ewald B. Nyquist

Associate Commissioner for Elementary, Secondary and Continuing Education
Walter Crewson

Associate Commissioner for Research and Evaluation
Lorne H. Woollatt

Assistant Commissioner for Research and Evaluation
William D. Firman

Assistant Commissioner for Occupational Education
Robert S. Seckendorf

Director, Division of Research
Carl E. Wedekind

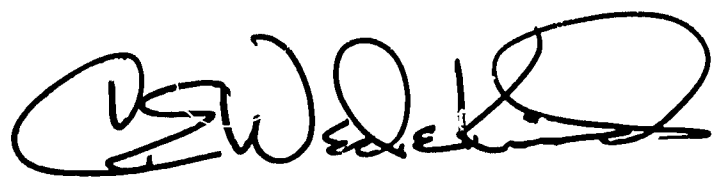
Chief, Bureau of Occupational Education Research
Louis A. Cohen

FOREWORD

The major purposes of this two year investigation were to determine the reliability and validity of the New York State written and performance competency examinations used in selecting candidates for teacher preparation in the trade and industrial programs. The three most widely used examinations, Auto Mechanics, Cosmetology and Machine Shop, were selected for investigation and further revision.

The study was jointly carried out under the auspices of the New York State Bureau of Occupational Education Research and the Division of Vocational Technical Education, College of Arts and Sciences, State University of New York at Oswego.

Educators wishing additional information regarding the specific instruments administered and scoring procedures followed, denoted in the text by appendix references, should contact the Bureau of Occupational Education Research, Room 574, New York State Education Department, Albany, New York 12224. Loan copies are available upon request.



Carl E. Wedekind
Director, Division of Research

ACKNOWLEDGEMENTS

Special thanks are due the Bureau of Occupational Education Research; Dr. Robert Seckendorf, Assistant Commissioner for Occupational Education, New York State Education Department; and the State University of New York for their support of this project.

The authors would like to thank Dr. Gordon McMahon, Director of Vocational Technical Education, State University College at Oswego, who conceived the idea for this research and arranged its funding.

Dr. Gordon Law, former member of the Division of Vocational Technical Education at Oswego, served as project administrator during 1966-67 and Dr. Edward Roden, former Syracuse regional Coordinator, has served as consultant throughout the project.

Thanks are due to Richard Thiele, Raymond Reisenger, Richard Walker and Dr. Leon Tunkel, (regional coordinators of the Vocational Technical Division at Oswego); Dr. Walter Zimmerman, Director of the Vocational Technical Division at State University College at Buffalo; Dr. Donn Billings and Veedar Colangelo of the State Education Department for their cooperation in arranging for and administering the tests.

Special assistance was received from Dr. C. Thomas Olivo, retired Director of Industrial Education, New York State Education Department.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

All schools with the major function of teacher preparation must recruit, select, motivate, educate and graduate persons who will become successful teachers. Enrolling and training competent persons for the teaching of occupational skills and knowledge into teacher education programs present special problems. Such persons must be able to teach effectively and must also have the skills and knowledge required in their occupations. Occupational competency is intimately related to the content of the programs in which such persons teach.

In order to accurately assess these skills and knowledge, reliable and valid measuring instruments must be used. The primary concern of this study is the problems associated with assessing the skill and knowledge level of the experienced tradesman who wishes to become a trade and industrial teacher.

There is a widespread need for vocational trade and industrial teachers. New York State has spent millions of dollars on building fine vocational education facilities. Staffing these facilities often becomes a real problem because of the difficulties involved in recruiting and training the required specialized teachers. The frequently cited need in our economy for skilled, rather than unskilled, workers implies that the need will increase for more people to teach the skills an expanding and changing economy requires. These teachers must not only instruct in the many newly developing fields, but must also be competent to teach many skills within a single occupation.

In many ways, the potential trade and industrial teacher is different from the prospective teacher in other fields. He is a mature person, usually in his 30's, who has had relevant work experience previous to entering the teaching field. He is leaving a career in industry or service occupations to become a classroom teacher.

Student Selection Procedures

The typical student selection procedures for the various trade and industrial teacher education programs in New York State are as follows, although the sequence differs somewhat at each of the teacher education centers:

1. Publicity
   a. The public is made aware of the existence of trade and industrial teacher education training programs.
   b. Newspaper stories are usually involved and also more informal methods for disseminating information about the programs are used by vocational education administrators and teachers.
   c. All interested persons are notified of date, time, and place for a meeting to further explain the program.
2. Explanation of the program
   a. Details of the program, qualifications, and type of expected employment are explained by the coordinator of the regional center to a large number of potential applicants.
   b. Applications are then accepted from interested tradesmen.
   c. Applicant qualifications are a valid high school diploma and five years of journeyman experience in the occupation.
3. Intelligence testing
   a. A group meeting is held for the applicants during which group intelligence tests are administered.
   b. The results of the tests are taken into consideration, particularly when the high school record of the applicant is questionable.
4. Interview
   a. Each applicant who meets the basic qualifications is individually interviewed, usually by the coordinator of the regional center and vocational technical administrator.
   b. The primary purpose of the interview is to assess the applicant's qualifications and potential as a teacher.
5. Written competency test
   a. A written competency test is administered to those applicants who have successfully completed the above steps in order to measure the degree of technical knowledge and understanding of the occupational field.
   b. There is a specific examination for each occupational field.

6. Performance competency test
   a. A performance competency test is given to those applicants who have passed the written competency examination to ascertain their actual occupational skills.
   b. These tests require the applicants to perform a number of typical tasks encountered in the occupation.
7. Final evaluation
   a. The regional coordinator evaluates all the information regarding the applicant and decides either to (1) admit, (2) reject, (3) admit provisionally, or (4) admit after the applicant has demonstrated increased proficiency by further work experience and then retaking and passing the occupational examination.

It is possible that the present selection procedures eliminate from consideration some tradesmen who could become competent teachers because the qualifications are not being assessed accurately. Or, on the other hand, students may be admitted who do not actually meet the qualifications.

Basic Questions Investigated

Two basic questions which this study was designed to investigate are (1) the reliability of the written and performance examination; i.e., do the examinations yield scores that are relatively stable for any individual; and (2) the validity of the written and performance examinations, i.e., do the examinations reflect the areas in which the applicant should possess skill or knowledge (content validity) and do the examinations differentiate the applicant's degree of skill or knowledge.

The proficiency examinations presently in use were developed statewide over a period of years and were essentially "teacher-made." The questions were written by trade and industrial teachers, and it was assumed that these tests were reliable and valid prior to the present study. The limited data collected on these tests concerned the percentage of applicants who scored above some arbitrary point, the observation that more applicants would "pass" the performance examination as compared to the written examination, and

scattered data on the level of difficulty of items for some of the written examinations. Many questions and criticisms have been raised by the people involved in using and taking these examinations. This study attempted to answer some of these questions and reduce the criticisms made of these examinations. An important part of this investigation concerned the use of reliability and validity data to review examinations and then attempt to improve the testing program.

Related Research

The use of proficiency examinations is widespread in the United States. Not only are proficiency examinations used for selection of vocational technical teachers in many states, but similar examinations are used for licensing workers in certain fields, and by the U.S. Employment Service for use where an applicant claims to know a particular job. However, there is not much evidence concerning the reliability or validity of competency examinations. Stead (1940) reports that certain trade tests were developed by testing experts, beginners and workers in related fields, retaining only those items that were able to discriminate between these groups. Burtt (1940) reports similar information for a trade test for engine-lathe operators. The small amount of evidence available supports the feasibility of obtaining the empirical data required for developing useful selection instruments.

There has been a recent attempt to begin working on some of the related problems in the area of selecting appropriate applicants for vocational technical teaching. Kynard's (1960) doctoral dissertation was concerned with the criteria for determining the value of work experience for teachers of trade and vocational education.

In recognizing many of the problems related to the examinations, the New York State Education Department sponsored a national conference at Rutgers University which dealt indirectly with the problems associated with the development of valid and reliable competency examinations. To emphasize some of the problems, one reported piece of research stated, "No attempt has been made to treat the information statistically. To do so would be an exercise in futility in that we are not sure enough of the reliability of the data." (LaBounty. 1967).

Proposals have been made to investigate the reliability and validity of competency examinations (Impellitteri, 1967), but little work has been done because of the many inherent problems involved in such investigations.

The consensus among those attending the conference was that there was an unquestioned need for reliable and valid competency examinations. (Griess, 1967).

METHODOLOGY

Written and performance competency examinations have been developed and used in New York State for a number of years. The three most widely used examinations, Auto Mechanics, Cosmetology and Machine Shop, were selected for investigation and further revision since it was felt that for these examinations there was a sufficient sample of applicants for accurate reliability and validity data. These examinations were written or revised during the summer of 1966 (1966 examination) by trade and industrial teachers under the direction of an area director of vocational education and with limited consultation from two specialists in psychological measurement.

The format of the 1966 editions of the written examinations required the examinee to record his answers directly on the test booklet. Each test booklet was separately scored by the examiner. In order to analyze the 1966 data, all information was transferred to IBM answer sheets (essay, completions and matching questions were coded on a correct-incorrect basis). This information was then put into a computer which item analyzed the results.

These examinations followed the "Guide for Preparing Occupational Competency Examinations for Admission to Industrial Teacher Training" published by the New York State Education Department, Bureau of Vocational Curriculum Development and Industrial Teacher Training. Although reliability and validity data was secured for all examinations that could be transferred to IBM answer sheets, the three exams cited above were the ones studied in detail.

Procedures for Standardization and Revision of Examinations

The first step in gathering the necessary data for assessing the reliability and validity of occupational competency examinations was an attempt to standardize procedures for administering and scoring the examinations. The manual for examiners was completely revised and expanded and it contained step by step instructions to the examiners on how to prepare to administer the examinations, as well as detailed procedures for the actual administration and scoring of them. (Re: Appendix A). Those sections of the manual pertaining to performance examinations were given special attention and clarified in minute detail. The two principal investigators travelled to each region of the state and met with the regional coordinators and prospective examiners in the fields of Auto Mechanics, Cosmetology and Machine Shop. The need for careful adherence to prescribed procedures, ways of handling problems encountered in administering the examinations and criteria to observe and rate during the examinations were discussed at these meetings. The goal was a statewide standard so that variations from center to center in administration procedures and scoring would be held to a minimum. The meetings were also useful for soliciting criticisms and suggestions for improving the examinations.

After meeting with the regional coordinators of the vocational teacher education program to assist in planning and scheduling the testing program, and to secure agreement on other steps in the screening process, the two principal investigators observed performance examinations at several centers. This was done so that problems in administering the examinations could be assessed, and also to identify any discrepancies in procedure. Other aspects of the screening process were observed, such as the interviewing of applicants. The later work was done primarily to acquaint the investigators with what was happening during other aspects of the screening process.

The examiners administering the tests at the various regional teacher education centers scored the written and performance examinations. Means and standard deviations were computed for those examinations that were given to ten or more applicants. Although this was part of the data-gathering phase of the study, it was also an important part of the administration of the testing program, since the scores and the normative data were immediately returned to the regional coordinators. On the basis of this information

decisions were reached on who passed, failed or would be retested on the examinations.

REVISION OF EXAMINATIONS: SUMMER 1967

Test Item Pool

In planning for a revision of the examinations, a collection of test items to be used for revising examinations was secured on a statewide basis. This was done by asking regional coordinators of vocational teacher education to contact directors of vocational technical programs in their geographical areas and to have these directors ask trade and industrial teachers from the appropriate fields to submit items. The vocational technical teachers were sent a brief guide prepared by the principal investigators for constructing objective test items. (Re: Appendix B). Hundreds of items were acquired through this procedure for the Auto Mechanics, Cosmetology and Machine Shop examinations.

To eliminate one source of error, a less cumbersome scoring procedure for the written and performance examinations was devised.

The 1967 editions of the examinations were designed to be machine scored as soon as they were received from the regional coordinators. IBM equipment was used for their scoring and for obtaining item analysis data. The item analyses of the examinations were used as a guide in determining necessary revisions of the examinations, as well as for research purposes.

After the first year of data gathering, the suggestions of examiners and others in the field and the item analysis data (when available) formed a basis for revising a number of examinations by a group of specialists in each occupation. Each group of specialists worked with the principal investigators for one week periods during the summer of 1967 revising all examinations materials in the fields of Auto Mechanics, Machine Shop, Cosmetology and the Printing and Carpentry examinations. This included analysis and restatement of the scope of each field, revision of both written and performance examinations, much more explicit directions to both the applicant and the examiner, and the development of new scoring procedures and rating scales for the performance examinations. (Re:Appendix C). Unfortunately, the item analysis and reliability data were not available for the Cosmetology examination until after the specialists had begun their work and

could not be used to identify items needing revision or elimination. In revising the examination, the groups of consultants carefully restated the scopes of the fields and determined in what areas the examinations needed revision. The item pool was extensively used so that it was not necessary for the specialists to spend time writing items, but rather to select what they considered to be good items from those that had been solicited.

In addition to the work on the major examinations the Printing and Carpentry examinations were revised since they had been heavily criticized. The Printing examination was reorganized into one examination in General Printing and one examination in Offset Lithography. Revised materials were also prepared for the Carpentry examination. Draftsmen were employed to draw sketches and illustrations used on all the revised examinations. A total of thirty persons worked with the principal investigators during the summer of 1967. In addition to the six examinations discussed above, the principal investigators also carried out minor revisions of five other examinations so that by the fall of 1967 the written examinations in the eleven most commonly tested fields were available in IBM machine-scoring format.

Evaluation forms were given to examiners throughout the State, and any criticisms and suggestions for improving the examinations were recorded for use in future revisions of the examinations.

Procedures for the Evaluation of Examinations

To assess the reliability of the written examinations, an item analysis of each of the examinations was performed. These data were used in revising the examinations in the summer of 1967. Analyses were also performed for the 1967 examinations and the eleven examinations that had been converted to IBM machine-scoring format.

Among other information, the item analysis data included the difficulty index (the percent passing each item) and the point biserial coefficient (a correlation of the item with the total test score).

Reliability and item analysis data for the 1967 examination and the 1966 examination were compared to determine whether or not the examination had changed significantly as result of the revisions. This was done for the 1966 examinations that had been converted to IBM machine-scored format in eight fields and for the 1966 and 1967 revisions of the examinations in Auto Mechanics, Machine Shop and Cosmetology.

One indicator of the reliability of an examination is the Kuder-Richardson formula # 20 which yields a measure of internal consistency, "If the items on a test have high

inter-correlations with each other and are measures of much the same attribute, the reliability coefficient will be high. If the inter-correlations are low, either because the item measures different attributes or because of the presence of error, then the reliability coefficient will be low." (Ferguson, 1966).

In order to assess inter-judge reliability on the performance examinations, the Machine Shop examinations were administered to a total of 25 applicants with pairs of judges rating the performers independently. One of each pair was the official examiner, and a second judge (an observer) was instructed to make his observations and ratings completely on his own without interacting with the official examiner. The rating scales for each pair of judges were then compared to assess the extent of agreement on the ratings.

Written examinations were administered to applicants who met the basic qualifications for admission to the teacher education program. Performance examinations were administered to as many of the applicants who took the written examination as was possible, regardless of their score on the written examination. For the 1966 examinations, a total of 143 applicants took both the written and performance examinations and 173 took the performance examination only. For the 1967 edition, 100 applicants took both the written and performance examination and 166 took the performance examination only. These figures represent the total number of applicants tested in the State during the respective years.

Written examinations were administered to persons whose general trade background was such that it was predicted that they would not do as well as regular applicants. The tests were administered to special samples who differed in verbal facility to see if scores on the examinations reflected trade skill or general test-taking ability.

The 1966 written examinations were given to a large sample of vocational high school seniors in the respective fields. From this larger group a random sample of 100 students was drawn for each field that was more representative of the population distribution of the State.

The 1967 written examinations were given to 112 vocational high school seniors and 15 adult education students in Syracuse, New York and 62 academic high school seniors in the Fayetteville-Manlius school system.

Another group taking the Auto Mechanics and Machine Shop written examinations were college seniors majoring in Industrial Arts at Oswego. Their scores were compared to those of vocational technical candidates, and comparisons within the Industrial Arts sample were also made, to test whether the amount of job experience was reflected in written examination scores.

A contemplated part of this study was to compare applicants to other groups on the performance examinations, but after several attempts this part of the project had to be abandoned. This was partly due to the prohibitive cost of performance examinations for special groups, and also due to the fact that many of the special groups available, (e.g., high school students) would obviously not be applicants since they would be much younger. Thus it would be impossible to determine if the examiners judged them with same standards as they judged applicants to the teacher education program. Data secured under such circumstances would be open to many interpretations.

All applicants who took written examinations in 1967 were also administered the performance examination, in order to acquire data for this study. The correlation of the written examinations with the performance examinations was computed, as well as the correlations of all sub-scores on the performance examinations with each other and with the total score and overall rating.

RESULTS AND DISCUSSION

Reliability

One commonly used estimate of the reliability of a test is the Kuder-Richardson internal consistency reliability coefficient. (KR-20). These coefficients were computed for the 1966 and 1967 editions of the written examinations. (See Table I). A coefficient of .90 is usually considered minimum (Diederich, 1960) for a standardized test and the coefficients for these examinations fall short of this level ranging from .712 for Cosmetology and .891 for Auto Mechanics. The KR-20 formula gives an overall index of the degree of inter-relationship between all the items on the examination.

TEST DATA

Table I

	% Correct	Mean	No. Items	S.D.	Range	No. Examinees	K-R No. 20	Std. Error
Auto Mechanics								
Applicants								
'66 Exam	65.15	68.45	105	10.41	44-89	55	0.847	4.075
'67 Exam	62.57	60.69	97	12.58	26-84	62	0.891	4.138
High School Seniors (Voc.)								
'66 Exam	39.11	41.07	105	13.52	9-68	100	0.905	4.149
'67 Exam	44.45	43.12	97	13.86	24-71	32	0.903	4.304
Adult Education								
'66 Exam	42.49	41.22	97	10.08	27-55	9	0.833	4.120
I. A. College Seniors								
'67 Exam	41.03	39.80	97	10.26	23-67	35	0.831	4.216
Cosmetology								
Applicants								
'66 Exam	66.78	79.47	119	12.36	49-107	58	0.883	4.233
'67 Exam	75.01	90.10	120	7.10	71-103	60	0.712	3.861
High School Seniors (Voc.)								
'66 Exam	59.67	71.01	119	14.82	33-103	100	0.909	4.460
'67 Exam	67.42	80.90	120	9.66	58-98	47		
High School Seniors (Acad.)								
'67 Exam	42.92	51.50	120	10.98	18-72	29		
Machine Shop								
Applicants								
'66 Exam	66.01	52.15	79	9.49	30-72	40	0.863	3.515
'67 Exam	67.53	57.40	85	7.96	40-74	35	0.788	3.663
High School Seniors (Voc.)								
'66 Exam	50.54	39.93	79	8.92	22-61	100	0.819	3.791
'67 Exam	53.33	45.33	85	9.50	32-70	33	0.821	4.021
High School Seniors (Acad.)								
'67 Exam	29.22	24.84	85	5.89	12-35	25	0.560	3.911
Adult Education								
'67 Exam	36.07	30.66	85	8.64	19-41	6	0.810	3.765
I. A. College Seniors								
'67 Exam	48.41	41.15	85	10.23	25-63	26	0.851	3.952

Another estimate of the reliability of an examination is the standard error of estimate. The standard error of measurement is the standard deviation of a sample of scores for an individual around his true score. If a person were to be re-tested indefinitely, approximately 68 per cent of his scores would be within one standard error of measurement of the true score. The greater the ratio between the standard error of measurement and the standard deviation of the test, the more reliable are the test scores. For the written examinations in this study, the standard error is less than one half of the standard deviation for all the examinations indicating high reliability.

Other important reliability data on an examination are the difficulty levels and point biserials for each item. The difficulty index is the per cent answering the item correctly and the point biserial is the correlation of the item to the total test score. The test writers discarded some items from the 1966 examinations, kept some items unchanged, revised others, and added some completely new items. Table II presents the average and the standard deviation of the discrimination indexes and point biserial correlations for all items on the 1966 and 1967 examinations and for the different categories in which the items belong.

Comparisons can be made between these categories on the difficulty indexes and point biserial correlations for the three examinations.

Analyzing all items on the 1966 and 1967 examinations finds the mean difficulty of items to be between 62.57 and 75.01, which is within the standards considered acceptable by Diederich (1960) for examinations. Although the difficulty level of the three examinations seems to vary considerably, these differences cannot be tested statistically since there is no way of assessing whether the applicants for the different fields are equivalent in terms of the various factors that determine the score on these examinations. The mean point biserial is much lower for all the examinations than is usually recommended.

The variability of the difficulty levels is extremely high, suggesting that there may be many items that lie outside a desirable difficulty level, i.e., either too easy or too difficult. Items that everybody or nobody gets correct do not significantly add to the total score in trying to discriminate between high and low characteristics. Examination of test item statistics does not indicate any significant improvement or change from the 1966 to 1967 examinations. The actual frequency distribution of the test statistics for

DIFFICULTY AND POINT BISERIAL INDEXES

	N	Mean Diff. Index	S.D. Diff. Index	Mean Pt. Bis.	S.D. Pt. Bis.
Auto Mechanics					
Discarded Items	68	.633	.172	.225	.145
Repeated Items - '66 Exam	30	.696	.168	.309	.141
Repeated Items - '67 Exam	30	.646	.188	.287	.165
Revised Items - '66 Exam	7	.701	.242	.189	.081
Revised Items - '67 Exam	7	.648	.118	.257	.066
New Items	60	.612	.168	.293	.143
Total - '66 Exam	105	.656	.240	.246	.147
Total - '67 Exam	97	.625	.206	.291	.142
Cosmetology					
Discarded Items	97	.647	.256	.260	.207
Repeated Items - '66 Exam	17	.761	.202	.229	.143
Repeated Items - '67 Exam	17	.711	.261	.176	.213
Revised Items - '66 Exam	5	.728	.194	.192	.106
Revised Items - '67 Exam	5	.634	.391	.180	.105
New Items	98	.754	.234	.177	.177
Total - '66 Exam	119	.667	.251	.253	.161
Total - '67 Exam	120	.711	.261	.177	.177
Machine Shop					
Discarded Items	52	.613	.267	.335	.196
Repeated Items - '66 Exam	11	.769	.144	.267	.173
Repeated Items - '67 Exam	11	.748	.124	.212	.186
Revised Items - '66 Exam	16	.683	.195	.251	.186
Revised Items - '67 Exam	16	.631	.218	.163	.132
New Items	58	.689	.240	.234	.171
Total - '66 Exam	79	.645	.244	.308	.192
Total - '67 Exam	85	.686	.225	.218	.166

N = Number of items.

Table II

the 1967 examinations are in Table III. For example, it can be seen that the Cosmetology examination was especially poor in that 82 out of 120 items were either too difficult or too easy and that only 32 out of 120 items had point-biserial correlations at the .30 level or above. Even though the other examinations were not this poor, they too could be improved by eliminating items that are too easy or too difficult.

Table III

FREQUENCY DISTRIBUTION OF DIFFICULTY AND POINT BISERIAL INDEXES - '67 EXAM

	Auto Mechanics	Cosmetology	Machine Shop
Difficulty Indexes			
0-.39	16	16	7
.40-.79	55	38	47
.80-1.00	26	66	31
Point Biserial Indexes			
Below .09	8	44	23
.10-.29	41	44	36
.30 & above	48	32	26

REASONS FOR LACK OF IMPROVEMENT ON THE REVISED EXAMINATIONS

A number of reasons for the failure of the revised examinations to yield reliability data superior to the 1966 examinations can be cited. The examinations were converted to an IBM machine-scoring format so that all items had to be forced choice selection items, with one right answer out of not more than five alternatives. Completion and matching items from the 1966 examination had to be discarded or converted to such a format, and problems could be used only as multiple-choice items. Although the practical

advantages of a test using standard machine-scored IBM answer sheets are obvious, the loss of some items that had excellent difficulty and discrimination indexes cannot be overlooked.

Another factor in the failure of the new examinations to yield superior statistical results is the fact that the item writers were not always using a statistical approach. Only a limited amount of statistical data was available at the time the new examinations were constructed, and the test item writers used basically a content validation approach in revising the examinations. The results appear to indicate the need for a more systematic statistical approach in revising the examinations in the future.

INTER-JUDGE RELIABILITY

A test of reliability of ratings on the performance examination in Machine Shop was conducted. Comparison between two judges who were independently rating the applicants indicates a high relationship between the two ratings. (See Table V). This would indicate that either the system of grading is not subject to interpretation bias of the examiner, or that the same biases operate for all examiners, or even that the ratings of the two judges were not independent. Although the two judges were told not to interact or discuss the work of the applicants, both judges were present in the same room and there might have been some subtle way in which they unwittingly exchanged information.

The intercorrelations between the various sub-scores on all the performance examinations reflect some interesting patterns. Each examiner is expected to observe various tasks performed by the applicant, rate each task on four dimensions, (skill, quality, speed and work habits), and give an overall rating. For the 1966 performance examinations, all of these intercorrelations were significant and very high.

An exact interpretation of the high correlations is difficult, since there could be many possible reasons why they exist.

Table IV

RELIABILITY AND VALIDITY DATA
ON OTHER MACHINE-SCORED EXAMS

	Mean	S.D.	Range	K-R No.20	Std. Error	Sample Size
Auto Body Repair - 97 items						
Applicants						
Fall '67	71.0	3.91	64-76	.24	3.40	7
Spring '68	73.0	3.38	67-77	.11	3.19	8
Vocational High School Seniors	59.9	8.96	46-75	.79	4.08	17
Carpentry - 153 items						
Applicants						
Fall '67	96.8	16.7	62-118			19
Spring '68	98.9	11.38	74-112			20
Adult Education Classes	58.3	13.00	44-81			9
Vocational High School Seniors	52.5	12.30	21-75			35
Electronic Servicing - 115 items						
Applicants						
Fall '67	68.6	15.4	43-96	.92	4.39	12
Spring '68	63.5	11.9	44-86			6
Adult Education Classes	35.0	11.34	12-44			6
Vocational High School Seniors	50.6	6.64	45-60			3
Food Service Trades - 97 items						
Applicants						
Fall '67	79.2	5.65	67-87	.66	3.29	10
Spring '68			-86			1
Vocational High School Seniors	67.7	5.33	58-76	.50	3.77	10
Academic High School Seniors	56.2	3.60	29-73	.76	4.18	33
General Printing - 105 items						
Applicants						
Fall '67	75.9	4.88	67-83	.49	3.49	9
Spring '68	77.00	10.31	60-86	.90	3.18	5
Vocational High School Seniors	73.7	15.39	56-83			3
Offset Printing - 105 items						
Applicants						
Fall '67	82.7	2.88	81-86	-.23	3.20	3
Spring '68	82.0	6.48	76-89	.75	3.21	4
Refrigeration & Air Conditioning - 98 items						
Applicants						
Fall '67	78.0	7.54	71-86	.87	2.70	3
Spring '68			-74			1
Adult Education Classes	62.6	11.43	42-80	.88	3.99	29
Commercial & Advertising Art - 100 items						
Applicants						
Fall '67						
Spring '68	78.4	11.63	48-93	.93	3.15	18

Table V

INTERJUDGE RELIABILITY OF MACHINE SHOP PERFORMANCE RATINGS

Examiner - Observer	=	.948
Mean rating by examiner	=	15.64
Mean rating by observer	=	15.42
Absolute mean difference between examiner and observer	=	1.34
S.D. of absolute mean difference	=	1.26

(1) It is possible that a good workman is competent in all aspects of the field and the poor workman is incompetent. If this were the case, it would be possible to administer a portion of the performance examination and get equally reliable results. (2) It is possible that since all the ratings are made by the same examiner, there is a consistent bias on the part of the examiner. For example, he might quickly form an opinion about the skill of the applicant based on observing one task and then consistently rate him high, average or low on all tasks.

To test the possibility that the "halo effect" was an important contribution to the high intercorrelations, comparisons were made among different types of scoring procedures. For the Machine Shop and Cosmetology examinations a revised scoring procedure was used, (Re: Appendix B), while for the Auto Mechanics examination the same scoring procedure was used. In addition, for the Machine Shop examination, two judges independently rated each applicant. There was no significant change in the intercorrelation matrix for performance scores between the 1966 and 1967 examinations. The new Cosmetology examination is constructed so that the examiner concentrates on the job the applicant is performing and this seemed to reduce the intercorrelations and lead to more independent judgements. For the Machine Shop

Table VI

CORRELATION MATRIX FOR AUTO MECHANICS - 1966 EXAM

	1	2	3	4	5	6	7	8	9	10	11	12	13
2	33*												
3	n.s.	78											
4	33*	82	68										
5	n.s.	56	36	60									
6	32*	70	54	73	68								
7	n.s.	71	54	68	82	71							
8	n.s.	61	50	53	69	45	73						
9	n.s.	58	35	55	79	54	70	71					
10	28*	92	79	84	65	78	79	72	66				
11	n.s.	81	68	80	55	78	72	58	47	83			
12	n.s.	92	80	85	56	72	74	67	59	94	86		
13	32*	76	72	73	46	71	66	49	45	73	63	80	
14	n.s.	82	68	82	88	79	91	82	84	90	78	85	71

Variables

1 - Written raw score
2 - Total performance
3 - Job No. 1 - Test readings
4 - Job No. 2 - Measurements
5 - Job No. 3 - Brake job
6 - Job No. 4 - Fuel pump test
7 - Job No. 5 -Trouble shooting
8 - Job No. 6 - Front end alignment
9 - Job No. 7 - Differential
10 - Skill for all jobs
11 - Speed for all jobs
12 - Quality of all jobs
13 - Work Habits for all jobs
14 - Sum total for all jobs

N	52 both written and performance
N	65 performance only
n.s.	not significant
*	significant at .05 level
rest	signigicant at .01 level

Table VII

CORRELATION MATRIX FOR COSMETOLOGY - 1966 EXAM

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	38														
3	30*	72													
4	28*	69	43												
5	n.s.	73	57	44											
6	33*	81	59	73	70										
7	n.s.	84	73	69	63	81									
8	44	80	67	61	61	70	80								
9	n.s.	61			41	67	66	54							
10	n.s.	69	69	61	52	49	64	69	n.s.						
11	40	78	76	65	56	68	75	72	68	64					
12	34*	88	78	76	72	83	90	83	79	76	85				
13	30*	83	70	75	70	82	81	72	69	62	69	85			
14	36	90	77	78	69	84	91	85	81	79	85	97	84		
15	33*	79	72	69	69	80	82	81	72	65	80	82	69	84	
16	36	91	80	80	74	88	92	86	79	81	86	97	89	97	91

Variables

1 - Written raw score
2 - Total performance
3 - Job No. 1 - Fingerwave
4 - Job No. 2 - Hair style
5 - Job No. 3 - Shampoo
6 - Job No. 4 - Hair cutting
7 - Job No. 5 - Permanent waving
8 - Job No. 6 - Hair coloring
9 - Job No. 7 - Wiggery
10 - Job No. 8 - Scalp massage
11 - Job No. 9 - Manicure
12 - Skill for all jobs
13 - Speed for all jobs
14 - Quality of all jobs
15 - Work habits for all jobs
16 - Total sum of all ratings

N = 60 both written and performance
N = 72 performance only
n.s. = not significant
* = significant at .05 level
rest = significant at .01 level

Table VIII

CORRELATION MATRIX FOR MACHINE SHOP - 1966 EXAM

	1	2	3	4	5	6	7	8	9	10	11	12
2	n.s.											
3	n.s.	90										
4	n.s.	71	84									
5	54	88	69	64								
6	n.s.	75	63	49*	67							
7	n.s.	83	80	71	73	71						
8	n.s.	94	89	76	86	78	90					
9	n.s.	84	83	68	85	84	82	87				
10	n.s.	89	90	70	79	81	83	90	83			
11	n.s.	86	93	85	69	68	75	83	78	84		
12	n.s.	94	94	82	88	84	90	96	92	95	92	

Variables

1 - Written raw score
2 - Total performance
3 - Job No. 1 - Grid 3 types of cutting tools
4 - Job No. 2 - Sharpen a twist drill
5 - Job No. 3 - Machining operations
6 - Job No. 4 - Set up milling machine
7 - Job No. 5 - Surface grid
8 - Skill for all jobs
9 - Quality for all jobs
10 - Speed for all jobs
11 - Work habits for all jobs
12 - Total sum of all ratings

N = 31 both written and performance
N = 36 performance only
n.s. = not significant
* = significant at .05 level
rest = significant at .01 level

Table IX

CORRELATION MATRIX FOR AUTO MECHANICS - 1967 EXAM

	1	2	3	4	5	6	7	8	9	10	11	12
2	31*											
3	34*	80										
4	n.s.	85	63									
5	n.s.	66	62	59								
6	n.s.	74	65	56	47							
7	n.s.	66	51	45	n.s.	56						
8	n.s.	81	56	67	54	44	53					
9	n.s.	58	34	40	35	45	30	35				
10	32*	98	79	82	66	72	64	81	57			
11	35*	97	80	84	67	70	60	81	57	95		
12	n.s.	83	62	64	46	59	53	71	54	81	79	
13	n.s.	86	67	78	61	62	55	71	37	84	83	75

Variables

1 - Written raw score
2 - Total performance
3 - Job No. 1 - Alternator
4 - Job No. 2 - Engine units
5 - Job No. 3 - Brake job
6 - Job No. 4 - Fuel pump
7 - Job No. 5 - Trouble shooting
8 - Job No. 6 - Front end alignment
9 - Job No. 7 - Wheel balance
10 - Procedure for all jobs
11 - Quality for all jobs
12 - Speed for all jobs
13 - Work habits for all jobs

N = 41 both written and performance
N = 64 performance only
n.s. = not significant
* = significant at .05 level
rest= significant at .01 level

Table X

CORRELATION MATRIX FOR COSMETOLOGY - 1967 EXAM

	1	2	3	4	5	6	7	8	9	10
2	n.s.									
3	n.s.	54								
4	n.s.	85	41							
5	n.s.	61	35	50						
6	n.s.	73	n.s.	44	29					
7	n.s.	72	50	47	48	44				
8	n.s.	92	53	70	62	63	64			
9	n.s.	84	48*	76	n.s.	n.s.	47*	n.s.		
10	n.s.	71	51	46	41	43	n.s.	71	n.s.	
11	n.s.	75	37*	68	65	n.s.	50	76	n.s.	n.s.

Variables

1 - Written raw score
2 - Total performance
3 - Job No. 1 - Fingerwave
4 - Job No. 2 - Hair style
5 - Job No. 3 - Shampoo
6 - Job No. 4 - Hair cut
7 - Job No. 5 - Scalp treatment
8 - Job No. 6 - Permanent waving
9 - Job No. 7 - Hair coloring
10 - Job No. 8 - Wiggery
11 - Job No. 9 - Manicure

N = 39 both written and performance
N = 72 performance only
n.s. = not significant
* = significant at .05 level
rest = significant at .o1 level

Table XI

CORRELATION MATRIX FOR MACHINE SHOP - 1967 EXAM

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	n.s.															
3	n.s.	91														
4	n.s.	79	77													
5	n.s.	65	74	39												
6	n.s.	87	83	71	46*											
7	n.s.	71	63	68	n.s.	46*										
8	n.s.	76	77	40	49	75	n.s.									
9	n.s.	99	92	79	62	87	69	76								
10	n.s.	99	90	78	66	85	71	75	98							
11	n.s.	76	76	96	n.s.	71	64	41	78	74						
12	n.s.	71	78	45	95	52	n.s.	58	69	72	42*					
13	n.s.	77	81	67	41*	88	44	63	80	72	65	42*				
14	n.s.	60	52	64	n.s.	42*	92	12	59	60	58	n.s.	n.s.			
15	n.s.	64	63	29	43*		n.s.	92	68	58	n.s.	46*	55	n.s.		
16	n.s.	76	72	96	39*	64		37	75	76	84	43*	62	64	n.s.	
17	n.s.	51	62	29	95	n.s.	n.s.	36	48	53	n.s.	80	n.s.	n.s.		n.s.

Variables

1 - Written raw score
2 - Total performance
3 - Overall rating
4 - Lathe-Procedure
5 - Milling-Procedure
6 - Surface grinding-Procedure
7 - Pedestal grinding-Procedure
8 - Drilling & Bendhwork-Procedure
9 - Observed performance total
10 - Finished product total
11 - Lathe-performance
12 - Milling-performance
13 - Surface grinding-performance
14 - Pedestal grinding-performance
15 - Drilling & Benchwork-performance
16 - Lathe-finished product
17 - Milling-finished product

N = 20 both written and performance
N = 30 performance only
n.s. = not significant
* = significant at .05 level
rest = significant at .01 level

examination the revised rating system requires the judge to rate the skill and finished product of the examinees. The intercorrelations between the various jobs are somewhat lower; however, there was almost perfect correlation between overall performance and overall quality of the finished product. This would again indicate that these two judgments were not independent. However, the evaluation of the examination should not be limited strictly to reliability statistics that might not indicate other ways in which the tests were improved, such as the selection of items consistent with the scope of the field.

Validity

The second important characteristic of a test is its validity, that is, does the test measure what it was designed to measure? There are a number of different ways of trying to determine the validity of a test, but the assessment of a test's validity requires a judgment and examination of the test itself.

The content or face validity of an examination is a frequently used method for both constructing and evaluating tests. This approach involves the use of qualified individuals determining whether or not the items appear to reflect important aspects of the knowledge or skill the test is designed to measure. For both the 1966 and 1967 examinations, a group of experts first discussed the scope of the examinations. They then analyzed the existing examinations to determine if certain areas were overemphasized or not represented. Items were then added or deleted on the basis of this analysis. Those persons responding to the questionnaires concerning the examinations felt that the 1966 examinations were an improvement over previous examinations and that 1967 examinations were better yet, since fewer questions were raised about these examinations.

Another way in which the face validity was improved was to eliminate items that reflected knowledge obtained in teaching the occupation, rather than knowledge that a person gains through experience. For example, the Cosmetology examination had many items dealing with the knowledge of Latin names for muscles. This type of information might be known by the occupational teacher since it is a part of the course content, but not by the person in the trade. It is possible for such an item to have a high correlation with the total test score and still not be a valid item.

Another method of determining validity is to see how the examinations are related to other criteria. Since these examinations are designed to differentiate the skill and knowledge learned in an occupation during the practice of this occupation, the scores on these examinations should be related to the number of years of experience the person has in the occupation. Since this data was not available on the applicants (All applicants had at least 5 years of experience), certain special samples were tested where it was known that the level of actual experience was less than that of the applicants. The scores of the applicants were significantly higher than those of the inexperienced samples. (See Tables XII and XIII).

Table XII

SIGNIFICANCE TESTS BETWEEN APPLICANTS AND VOCATIONAL HIGH SCHOOL STUDENTS - 1966 EXAMS

	$\bar{X}$	S.D.	N
Auto Mechanics			
1. Applicants	68.45	10.41	55
2. Vocational High School Students	41.07	13.52	100
Z 1-2 = 13.90 $p < .001$			
Cosmetology			
1. Applicants	79.47	12.36	58
2. Vocational High School Students	71.01	14.82	100
Z 1-2 = 3.52 $p < .001$			
Machine Shop			
1. Applicants	52.15	9.48	40
2. Vocational High School Students	39.93	8.92	100
Z 1-2 = 6.90 $p < .001$			

Table XIII

SIGNIFICANCE TESTS BETWEEN CANDIDATES AND

OTHER SAMPLES - 1967 EXAM

	$\bar{X}$	S.D.	N
Auto Mechanics			
1. Applicants	60.69	12.58	62
2. Vocational High School Students	43.12	13.88	32
3. Adult Education	41.22	10.08	9
4. I.A. College Students (all)	39.80	10.26	35
5. I.A. College Students (experience)	45.33	8.65	12
Z 1-2 = 5.92 $p < .001$			
t 1-3 = 4.98 $p < .001$ (d.f. = 69)			
Z 1-4 = 8.83 $p < .001$			
t 1-5 = 5.02 $p < .001$ (d.f. = 72)			
Cosmetology			
1. Applicants	90.10	7.10	60
2. Vocational High School Students	80.90	9.66	47
3. Academic High School Students	51.50	10.98	29
Z 1-2 = 5.41 $p < .001$			
Z 1-3 = 16.92 $p < .001$			
Machine Shop			
1. Applicants	57.40.	7.96	35
2. Vocational High School Students	45.33	9.50	33
3. Academic High School Students	24.84	5.89	25
4. Adult Education	30.66	8.64	6
5. I.A. College Students (all)	41.15	10.23	26
6. I.A. College Students (experience)	44.36	9.83	11
Z 1-2 = 5.56 $p < .001$			
Z 1-3 = 17.89 $p < .001$			
t 1-4 = 5.46 $p < .001$ (d.f. = 39)			
Z 1-5 = 6.58 $p < .001$			
t 1-6 = 3.84 $p < .001$ (d.f. = 44)			

Within the sample of Industrial Arts college seniors, it was possible to set up two separate groups, those with some type of experience related to the examination and those without this type of experience. (See Table XIV). The experienced Industrial Arts seniors were superior to the inexperienced seniors.

These results are particularly impressive since they refer to the written examination, where it was expected that there might not be any significant differences between the applicants and the other groups, especially the Industrial Arts college seniors. It was also expected that verbal ability would be an important biasing factor in these examinations

Table XIV

COMPARISON OF AMOUNT OF RELATED EXPERIENCE OF INDUSTRIAL ARTS STUDENTS TO WRITTEN EXAM PERFORMANCE

	Mean Score	S.D.	N
Auto Mechanics			
I. A. Students with job experience or related hobbies or club.	45.33	8.65	12
I.A. Students without job experience or related hobbies or club.	36.55	11.11	20
$t = 2.29$ $p < .05$			
Machine Shop			
I.A. Students with job experience or related hobbies or club.	44.36	9.83	11
I.A. Students without job experience or related hobbies or club.	35.00	14.73	17
$t = 1.79$ $p < .10$			

but a comparison of academic (high verbal ability) with the vocational high school students (low verbal ability) indicates that the relevant occupational training and not just verbal skill is what is being measured by the written examination.

Still another way to investigate the validity of examinations is to examine the relation of the examination to some other variables. In this investigation there were two types of competency examinations, a written test and a performance examination. If these tests are both measures of skill and/or knowledge gained with occupational experience, there should be a positive correlation between them. Column 1 of Tables VI to XI show the correlation of the written with the performance examination.

For both the old and new examinations it was requested that all applicants for admission to the vocational teacher education program who were administered the written examination also be given the corresponding performance examination. The resulting intercorrelation matrices indicate that for the 1966 examination there was a correlation between the written examination and the performance rating for the Cosmetology and Auto Mechanics examinations but not for Machine Shop. For the 1967 examinations only Auto Mechanics was related to the performance rating, which indicates that the written examination is measuring something different than is being measured by the performance examination. Apparently it is possible to differentiate between the knowledge of the area that the person has and the skill in the execution of his job.

However, the fact that these correlations are non-existent or very low, does not necessarily mean that skill and knowledge are unrelated. Before reaching such a conclusion, it must be determined what each examination is measuring and whether it measures it reliably. The lack of correlation between written and performance examinations indicates that they are measuring different factors, but it does not tell us what these factors are. It is possible that the factors measured by the written and performance examinations are unrelated because either or both of the examinations are not measuring anything related to trade knowledge or skill. Only by assuming that occupational knowledge is being measured by the written examination and that occupational skill is accurately reflected in the overall performance ratings can we conclude that these two facets of an occupation are unrelated or that

they are not measuring skill developed with experience.

RECOMMENDATIONS ON THE OCCUPATIONAL COMPETENCY TESTING PROGRAM

After studying the reliability and validity of several occupational competency examinations, and the problems inherent in the program, certain recommendations can be made that are consistent with standard testing procedures.

An obvious problem which faces a testing program of this type is the small number of applicants who take any one test each year. This makes it extremely difficult to establish norms and cutoff points, and to secure the necessary data to validate the tests. This problem is even more acute when one considers that the examinations must be kept up to date, so that using applicants over a number of years to achieve a large sample is ruled out. It is recommended that procedures throughout the State be completely standardized, and that data from all areas of the State be pooled for purposes of establishing norms and used in making decisions.

Along with the above recommendation, the most practical means for standardizing procedures and pooling data would be to send all examinations to a central point for scoring and item analysis, and then to disseminate the results to the regional coordinators. It is recommended that all written exams be converted to an IBM machine scoring format to help accomplish this. In this way, item analysis and reliability data can be secured routinely on every administration of the examinations, and distribution of scores, norms, and cutoffs can be established immediately on a statewide basis. This has been achieved for the eleven examinations discussed in this study, and it is suggested that this be the goal of all revisions in the future.

Considerable confusion, as well as problems in establishing norms and conducting test development research is caused when examinations are administered at different times through the State. It is suggested that the major testing program should take place once each year and that if written examinations would be administered in May and performance examinations in June all candidates could be notified of their

status before enrolling for courses in the fall.

Problems of standardizing the administration and scoring procedures are even more crucial when considering performance examinations. It is not enough to have qualified judges and good examinations--one must also assure that all precedures are the same, and other variables that might influence the score are the same at all the testing centers. One way to achieve this is to have training classes for examiners, bringing them together from all parts of the State, by field, instructing them in the administration of the examination, then certifying them as examiners in their region.

Another recommendation for standardizing the administration of the performance examination is that each one should be administered to as many candidates as possible at one time and in one place. For example, instead of testing four auto mechanics in Syracuse, four in Rochester, and four in Albany, all twelve could be tested at once and in one place by three judges. This would not only standardize the situation and procedures for all candidates, but would also allow a check on interjudge reliability. The ideal would be to have all auto mechanics tested at one center and all cosmetologists at another. Of course, the practical problems inherent in such a plan are recognized.

Written Examinations

The major recommendation relative to the written examinations is the development of a system of revising and up dating examinations on a planned and rotating basis. The use of statistical data and standard item analysis procedures has been demonstrated to be useful in such revisions, but it is also essential for this to be combined with content validation procedures. It is a questionable gain if an examination improves statistically but is seen as invalid by experts in the field.

One of the important needs at present, in relation to content validity, is the involvement of industry in the revision of these exams, Although attempts were made during the two years of this study to include on the test revision panels persons who were in no way involved in teaching, the examinations remain primarily teacher-oriented. It is recommended that any revision of examinations in the future include a systematic appraisal by non-teachers in the occupation for content validation purposes.

Test Item Pool

Another problem in revising written examinations is the need for a substantial pool of up-to-date items. Such a pool was obtained for the three examinations studied most intensely in this research (Machine Shop, Auto Mechanics and Cosmetology) by soliciting them from vocational technical teachers through the State. This procedure was quite successful and very economical, but it did not include items from persons who were outside of the field of vocational education. The writers feel that this means of securing an item pool should be continued in the future on a much broader basis, for all examinations, and that items should be solicited from non-teaching professionals also, even though this will probably require some payment for items accepted.

Although the data in this study indicates that the examinations do a reasonably good job of discriminating skilled and experienced tradesmen from those with inadequate skill and experience, it is felt that further improvement of these exams requires a clearer statement of the objectives of the competency examinations. The present use and interpretation of the examination scores implies at least four major objectives for these tests; a) to measure experience at the trade, b) to measure actual trade skill, c) to predict future success as a teacher and d) to give college credit. Any one of these objectives would present a problem of sufficient complexity to keep a measurement expert busy for years. An attempt to combine all four objectives without determining which is the major goal of the tests is not likely to measure any one of them successfully.

Another serious problem is that frequently the job experience a skilled person has is not necessarily the type of experience needed to qualify for a teaching position. Often he will specialize and have considerable experience in just one aspect of the field in which he is expecially proficient. The decision as to whether the test is to reflect occupational competency as the occupation is currently practiced or a rather unusual, broad experience, must be made.

Further Research

This study has demonstrated the feasibility of using statistical item analysis and cross validation procedures to assess and improve the effectiveness of occupational competency

testing programs, and it has also highlighted the need for further research in a number of areas.

For example:

a) Inter-judge reliability studies, such as that conducted on the Machine Shop performance examination, should be designed and carried out on all performance examinations on a regular basis.

b) A careful study of the characteristics of the applicant population, including data on those accepted and those rejected. should be reported on a yearly basis.

c) The predictive validity of these examinations and of other steps in the screening process should be carefully explored.

d) Tradesmen in the various fields, who are not applicants for teacher education, should be tested as a further cross validation procedure.

e) Content validation studies, employing tradesmen and other industry representatives as well as leading educators, should be initiated.

f) Research on new approaches to scoring performance examinations, which was begun on a few of the examinations this year, should be continued and expanded.

g) The relationship between both type and quanitity of experience, on the one hand, and performance on these examinations, should be explored.

h) The characteristics of successful trade and industrial teachers should be empirically identified.

Needs for a Testing Office

In order to begin to carry out the recommendations previously stated, it is suggested that a testing office be established to carry out test administration, scoring, reporting and research and development on a statewide basis.

The many and varied problems encountered in administering a statewide written and performance testing program in a wide range of occupations require the direction and coordination of a single central office. This office should be assigned the occupational competency testing program as its only function, since this assignment would provide sufficient important and complexity for any one office. This office, initially, should

be staffed by a full time director of testing, at least one full time secretary, and a budget which would allow for considerable supplies, consultant fees and extra clerical help. It also requires office space which allows for storage of all testing materials and data dating back several years, and accumulating each year.

The four major functions of such a testing office would be: (a) test administration, (b) test scoring, (c) test reporting, and (d) research and development. Each of these functions is crucial to the success of occupational competency testing, and statewide control in each of these areas is essential. If New York State could provide such services, it is probable that other states would want to participate in the testing program on a fee basis, thus greatly expanding the size of samples and providing some financial support.

Typical functions of this office would be as follows:

A) Supervising the revision and up dating of examinations on a planned, rotating yearly basis.

B) Assessing the need for examinations in new occupations, and supervising the writing of these examinations.

C) Providing for the training of performance examiners, and continually working toward standardized procedures throughout the State.

D) Working with regional coordinators in setting examination dates, arranging for examinations, deciding on policies.

E) Distributing tests, materials and information to the regional coordinators.

F) Providing test scoring and analysis services, and using statewide statistical and normative data to aid regional coordinators in making acceptance and rejection decisions.

G) Continually working toward maximum utility of occupational competency examinations is the total teacher selection process.

H) Specifying needed areas of research and cooperation with research personnel in carrying out such research.

I) Working with the State Education Department toward a long range plan for objectives, priorities, and directions for the occupational competency testing program.

The above is not meant to be an exhaustive list of the functions of such a testing office, nor is it meant to limit the Director of Testing who would be in charge of sucha

facility. A highly qualified person would need to develop the goals and functions of such an operation to fit his perceptions as he studied the statewide situation and learned of the problems to be solved. The Director would also want to integrate his efforts with the appropriate departments of the State Education Department and the State University. Such an office is desperately needed, and these are the types of functions that would be served.

SUMMARY

The major purpose of this investigation was to determine the reliability and validity of written and performance competency examinations used in selecting candidates for teacher preparation in trade and industrial education programs. One of the important criteria for successful teaching in the trade and industrial area is occupational competency related to the field in which the person is going to teach. In order to enter the field of vocational trade and industrial education the person must, according to New York State law, have five years of journeyman (or its equivalent) experience. In addition to having the relevant work experience the person must pass trade examinations in his particular field as additional evidence that the person is competent in his field, and also meet other selection criteria. It is possible that the present selection procedures eliminate from consideration some tradesmen who could become competent teachers or admit students who do not actually meet the criteria because trade skills are not being accurately assessed by the competency examinations.

Of the various steps in the selection procedures for entrance into vocational teaching, the testing program was singled out for careful investigation. Although competency examinations have been in widespread use throughout the United States for many years, they have not been subject to the same careful analysis that other examinations have been. The problem of accurately assessing the reliability and validity of competency examinations has been recognized to be time consuming and expensive. For these reasons, very little work has been done in this area, although this type of examination is constantly receiving a great deal of criticism from those who administer, take, or have to interpret the results of such examinations.

The proficiency examinations used in New York State have been written by teachers from vocational high schools and were revised when they appeared out-of-date or invalid because of criticisms from various sources. The initial phase of this study involved obtaining the basic statistics for analyzing the reliability and validity of the examinations being used. A basic attempt was made to improve the reliability of the examinations by familiarizing the examiners of the written, and especially of the performance examinations, with the specific duties involved in the administration of examinations. The was done to cut down on the amount of variability in actual testing procedures that existed between the various testing centers.

The format of the 1966 editions of the written examinations required the examinee to record his answers directly onto the test booklet. Each test booklet was then separately scored by the examiner. In order to analyze the 1966 data, all information was transfered to IBM answer sheets. The answer sheets were then re-scored by a computer which also item analyzed the results.

The format of the 1966 edition of the performance examinations had each candidate perform a number of different tasks typical of the field. Each task was rated by the examiner on four different dimensions; skill, speed, quality and work habits and the examiner also gave an overall rating for the total performance examination. The written examination score, the total performance rating, and the sub-ratings on the performance tasks were punched onto IBM cards for each candidate in order to compute the correlation between written and performance competency and the inter-relation among the various tasks used on the performance examination.

A state-wide random sample of high school students was also given the written examination, since they have less trade experience than the applicants of the teacher training program.

The results of the administration of the 1966 edition for the Cosmetology, Machine Shop and Auto Mechanics examinations indicated:

a) The Kuder-Richardson reliability coefficients were adequate.

b) The point biserial item correlations were moderate.

c) The level of difficulty between the examinations varied considerably.

d) There was only a slight or zero correlation between written and performance examinations.

e) There was an extremely high correlation between total performance rating and the rating for specific jobs on the performance examination.

f) There was confusion in the scopes of the examinations and the directions to the applicants and in the directions to the examiner.

g) High school students scored significantly lower on the written examinations than did the applicants.

Although the above results were somewhat encouraging, an attempt was made to see if it was feasible to improve upon these examinations. One major difficulty in developing an examination is writing items that are clear and unambigious, but differentiate between good and poor applicants. So that the people revising the examinations could concentrate their efforts towards clearly defining the scope of the field rather then spend time in writing items, regional coordinators solicited a large number of items from vocational trade and industrial teachers in the fields of Cosmetology, Machine Shop, and Auto Mechanics. These (and other examinations which were heavily criticized) were revised during the summer of 1967 by selected vocational teachers. The written and performance examinations for each of these fields was revised. The written examinations were made completely machine scorable and the performance rating system was revised for the Cosmetology and Machine Shop examinations.

The revised examinations were administered to applicants and other groups selected for comparison purposes. Special consideration was given to trying to assess the reliability of the judgments made on the performance examinations. For the Machine Shop examination there was an official examiner and a separate observer who also made ratings on the candidates performance.

Analysis of the 1967 data indicated that it is possible to improve the examinations and the examination procedures. This was indicated by:

a) Increased reliability coefficients for some examinations.

b) High interjudge reliability for the Machine Shop performance examination.

c) Strong indications of validity, in that the examinations differentiated between levels of experience.

d) Reduced correlations between total performance rating and separate job ratings.

e) Fewer complaints were received concerning the scopes, the directions to examiners or the examinations themselves.

Conclusions were drawn concerning:

a) The feasibility and need for further revision and improvement of examinations.

b) The need for the establishment of a center for conducting the testing program for competency examinations and suggestions for the operation of such an office.

c) The need for further research concerning these examinations and other major questions in the field of vocational technical education.

BIBLIOGRAPHY

Burtt, Harold E. Principles of Employment Psychology (2nd edition). New York, Harper 1942.

Diederich, Paul B. Short-Cut Statistics for Teacher-Made Tests. Princeton, New Jersey Educational Testing Service, 1960.

Ferguson, George A. Statistical Analysis in Psychology and Education (2nd edition). New York, McGraw-Hill, 1966.

Greiss, Gerald A. Feasibility of Providing Trade Competency Examinations for Teachers on a National Basis. Washington, U.S. Department of Health, Education, and Welfare,

Impellitteri, Joseph T. "Constructing Valid Occupational Competency Examinations,"in Feasibility of Providing Trade Competency Examinations for Teachers on a National Basis. Washington, U.S. Department of Health, Education, and Welfare, 1967.

Kynard, Alfret Tennyson. "Criteria for Determining the Value of Work Experience for Teachers of Trade and Industrial Education," University of California, Berkeley, Unpublished Dissertation, 1960.

LaBounty, Ray A. "A Limited Field Test of Automotive Competency Examination," in Feasibility of Providing Trade Competency Examinations for Teachers on a National Basis. Washington, U. S. Department of Health, Education, and Welfare, 1967.

Stead, William H. and others. Occupational Counseling Techniques. New York, American Book, 1940.