Occupational licensure: testing as a tool for measuring minimum competency

Janice Marie Turner
Atlanta University

Follow this and additional works at: http://digitalcommons.auctr.edu/dissertations
Part of the Public Affairs, Public Policy and Public Administration Commons

Recommended Citation
OCCUPATIONAL LICENSURE: TESTING AS A TOOL FOR MEASURING MINIMUM COMPETENCY

A DEGREE PAPER
SUBMITTED TO THE FACULTY OF ATLANTA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PUBLIC ADMINISTRATION

BY
JANICE MARIE TURNER

DEPARTMENT OF PUBLIC ADMINISTRATION

ATLANTA, GEORGIA
JULY 1988
ABSTRACT

PUBLIC ADMINISTRATION

TURNER, JANICE MARIE B.A., Spelman College, 1985

Occupational Licensure: Testing as a Tool for Measuring Minimum Competency

Adviser: Dr. Keith Simmonds

Degree Paper dated July 1988

The purpose of this paper is to examine the effectiveness of testing as a tool for measuring minimum competency as it applies to occupational licensure. The primary purpose of State Examining Boards for licensure and certification is to protect the public from unsafe or incompetent practice of individuals in a profession or trade by which the public may suffer harm through loss of life, health or property rights. The significance of this study is to examine the impact testing has in attempting to assess an individual's level of minimum competency for occupational licensure.

In the current debate on occupational licensure, the following questions have been raised: How accurately can minimum competency be measured? What will be the standards for measuring minimum competency? Is a passing score on a State Board examination an accurate reflection of minimum competency? In an attempt to address these concerns the writer identified four areas of licensure testing that impact greatly on the credentialing process. These areas included: test validity,
test reliability, test bias, and test policies. The main sources of information were informal conversational interviews and participant observation. Secondary sources included books, articles, lectures and journals.

While the writer's findings were inconclusive, and many of the problems that are common with testing were not evident with the State of Georgia Examining Boards Division, Examination Development and Test Administration Section the writer offered several recommendations in an effort to improve on the efficiency of the Examination Section of the Board.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>iii</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. THE PROBLEM AND ITS SETTING</td>
<td>4</td>
</tr>
<tr>
<td>III. REVIEW OF THE LITERATURE</td>
<td>8</td>
</tr>
<tr>
<td>IV. METHODOLOGY</td>
<td>14</td>
</tr>
<tr>
<td>V. ANALYSIS</td>
<td>16</td>
</tr>
<tr>
<td>VI. CONCLUSION AND RECOMMENDATIONS</td>
<td>34</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>36</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>42</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Basic Types of Validity</td>
<td>20</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

The primary purpose of State Examining Boards for licensure and certification is to protect the public from unsafe or incompetent practice of individuals in a profession or trade by which the public may suffer harm through loss of life, health or property rights. Medical licensure laws date back to the early part of the 19th century when they were enacted at the behest of State Medical Societies. Dr. Benjamin Shimberg, a noted authority on occupational licensure and certification, cites the purpose of medical licensure laws. He states that:

Without licensure, better trained physicians could gather together in medical societies, set high entry standards, promulgate codes of ethics, and even eject from membership those who were incompetent or who refused to adhere to these codes, but none of this could prevent quacks and charlatans from continuing to practice. The only solution seemed to be legislation that would make it illegal for anyone to practice without a license.¹

Examinations play a major role in ensuring that this purpose is served by the State Examining Boards. The examinations given for each profession or trade must accurately reflect the minimum standards for safe and competent practice in that field. Minimum standards are the essential knowledge, skills and abilities an individual must be able to demonstrate in order to ensure that the public health and welfare will

be protected if that individual practices in his/her chosen field. Examination standards should require no more than these essential attributes and certainly, no less. To require more may impose unfair limitations on the free trading of goods and services in our society and the rights of the individual to practice in his/her chosen profession. To require less, is to neglect the charge given by the State to protect its citizens from harm.

In examining for fitness for licensure or certification, a delicate balance must be struck between the protection of the public health and welfare and the protection of an individual's rights to pursue the profession or trade of his/her choice. Moreover, each board member has a responsibility for ensuring that Board rules, regulations, policies and practices are consistent with the general principles of fairness and equal opportunity for all persons coming before them for licensure. Every step in the process must be directed toward the dual goals of guaranteeing fairness and ensuring standards for minimum competency.

It is important to distinguish between licensure and certification. According to The U.S. Department of Health, and Human Services:

Licensure is defined as a process by which an agency of government grants permission to an individual to engage in a given occupation upon finding that the applicant has attain the minimal degree of competency required to ensure that the public health, safety and welfare will be reasonably well protected.²

Thus, licensing usually identifies the scope of the practice and contains statements that define what a licensed practitioner's duties and

²Ibid., p. 101.
responsibilities entail. Certification is a form of special acknowledgment:

Certification is the process by which governmental or non-governmental agency grants recognition to an individual who has met certain pre-determined qualifications set by a credentialing agency. Certification is sometimes referred to as a form of 'title control.' It is a way of identifying individuals who have met some standard. Unlike licensure, a certification law does not prohibit uncertified individuals from practicing their occupations. However, only those who have met the standard set by a governmental agency are permitted to use a designated title.

In current debate on occupational licensure, the following questions have been raised: How accurately can minimum competency be measured? What will be the standards for measuring minimum competency? Is a passing score on a State Board examination an accurate reflection of minimum competency? Indeed, these questions have particular relevance to this study. In an attempt to address these concerns, this paper presents an exploration of the problems associated with occupational licensure and the State Examining Board's responsibility to protect its citizens' health, safety, and welfare.

\[3\text{Ibid., p. 103.}\]
II. THE PROBLEM AND ITS SETTING

Background of the Agency

The writer interned with The Office of the Joint Secretary, State Examining Boards Division in Atlanta, Georgia. The State Examining Boards, a division of the Office of Secretary of State is provided for under the Official Code of Georgia Annotated, Title 43, Chapter 1.

The Executive Reorganization Act of 1931 created the Examining Boards as a Division of the Office of Secretary of State effective January 1, 1932. The 1931 Act abolished the office of secretary for each board and established the position of joint secretary as division director and executive administrative officer for boards. The purpose and intent of the reorganization were to bring regulatory boards under one central agency to give relief to the Executive Department. Prior to 1932, boards were administered separately and were accountable to the Governor for their actions.

The State Examining Boards Division of the Office of the Secretary of State is responsible for licensing individuals who are employed in various occupations in the State of Georgia. Depending on the legislation that created each board, individual board members are appointed by the Governor and approved by the Secretary of State and/or confirmed by the Senate. The number of individuals appointed to each board varies and the terms of appointment vary from three to six years. All persons appointed to the boards serve without compensation but do receive per
diem and reimbursement for travel expenses incurred in the execution of
the board's functions. The State Examining Board is assigned the admin-
istrative responsibilities for various licensing boards established by
the State of Georgia General Assembly. In addition, the State Examin-
ing Board also provides financial and investigative functions for the
boards under a centralized administrative structure in order to promote
efficiency, coordinate policy and to increase public accountability.
(For a more detailed composition of the Office of The Secretary of
State, the Office of The Joint Secretary of State, Examining Boards
Division and the organization of the State of Georgia, see Appendix A).

In general, the activities performed by all of the boards, to ful-
fill their role in regulating professions and businesses by individual
employees, are similar and include the following:

- Determining the requirements for licensure;
- Establishing the general scope of examinations;
- Issuing and renewing licenses to qualified persons;
- Adopting rules and regulations on how the profession
  should be conducted; and

Adopting and enforcing provisions of the law by:

- Authorizing investigations of complaints;
- Transmitting cases to the Board Attorney for hearings;
- Prosecuting cases in criminal court;
- Revoking, suspending or denying licensure applications; and
- Enjoining unlicensed persons from practicing a specific
  profession.

Located in the office where the intern worked is the Examination
Development and Test Administration Section. This Section is charged
with the responsibility of administering occupational examinations for
licensure and certification. The Examination Development and Test
Administration Section has been in existence only since 1982. This
Section was created in an effort to centralize the administration and advising of results to candidates for all licensing boards. The Examination Development and Test Administration Section consists of an Executive Director, two Examination Development Consultants, three Examination Technicians and a Support Staff Assistant.

**Internship Experience**

As an intern in the Examination Development and Test Administration Section, the writer worked closely with the Executive Director, Test Consultants and other Directors from various licensing boards. Consequently, the intern was exposed to the licensing process which included: application review, developing rules and regulations, disciplinary proceedings and other board related activities.

The writer's primary project was to develop a working module for the State of Georgia's Private Detectives and Security Agencies licensing examination. In an effort to prepare the document, the writer was assigned readings regarding test design, test construction, item development, item analysis, validation procedures, and job analysis.

During the course of the internship, the writer was exposed to and participated in various facets of the State Examining Board's activities such as: monitoring grievance hearings with various licensing boards and attending regular meetings with the Executive Director. In addition, the intern's other responsibilities included:

1. Assisting Examination Technicians with test preparation and test administration;
2. Serving as a model for the manicurist licensing examination;
3. Helping to proctor various examinations when the Examination Technicians experienced a shortage of support staff proctors.
Finally, the internship provided such a diverse experience that the writer's initiative allowed her the opportunity to explore other aspects of the agency independent of assigned duties and responsibilities. Other independent activities that the writer became involved with included: attending practical portions of examinations, and observing how decisions, disputes and discrepancies are settled in accordance with the State of Georgia's licensing policies and procedures.

Statement of the Problem

The problem to be addressed in this paper is two-fold and may be stated as follows: How dependable are Test Validity and Test Reliability as methods for measuring minimum competency as they apply to occupational licensure? In addition, to what degree do testing policies require that these examinations ensure that the minimal level of competency is being achieved and maintained? These questions are at the center of the controversy over testing for licensure. Minimum competency has been defined as the essential knowledge, skills and abilities an individual must be able to demonstrate in order to ensure that the public health and welfare will be well protected if that individual practices in his/her chosen field.

Occupations currently covered by some form of licensing or certification are many and diverse. However, the writer's focus is primarily on licensed occupations, although many of the issues and concerns identified earlier apply equally to alternative modes of credentialing - such as certification.
III. REVIEW OF THE LITERATURE

Standardized tests can be used to open or close doors to preparatory schools, colleges and professional school. Standardized test results can be used to decide who qualifies for government jobs or who will advance in business careers or ventures. These test results determine whether, after years of sweat and study, an individual is, indeed, ready to practice law, medicine, accounting, teaching, and many other licensed professions. But, whether one likes them or not, standardized test results are often considered to be the most important or at least, most objective measures of achievement, or potential suitability and there is no prospect of their disappearing anytime soon.

The focus in professional and occupational testing is on the level of knowledge and skills necessary to reasonably assure the public that individuals are competent to practice. The most fundamental challenge any licensing examination is validity. A state licensing agency is legally charged with establishing a testing standard that is acceptably valid for the purpose of ensuring functional competence. The instruments that are used to assess the minimum level of competency, must, in fact, reflect changes in the nation's curriculum - changes that are seen not only as a result of differing emphasis, but also changes that are more regional in nature.

In licensure testing the most important decision for which test results are used is to separate those applicants who deserve to be
licensed from those who do not. According to Dr. Shimberg,

Occupational licensing is a paradox. On one hand, licensing restricts the freedom of the individual, a concept we hold sacred in America. Yet licensing is intended to serve the overriding purpose of protecting the health and safety of the many.4

Licensure is the most restrictive type of statutory credentialing because it defines who may carry out certain activities and makes it illegal for anyone not licensed to do so. Dr. Shimberg states that:

If the protection of the public interest is the cornerstone on which the edifice of occupational licensing rests, the test used to determine competency are clearly one of the main pillars that support the structure itself. To the public, to legislators, to workers in occupations which are licensed, and to licensing board members themselves the test is the objective standard that separates the competent practitioner from the one who may not be trusted to function properly and safely.5

The goal in test development is to create items that discriminate between candidates who can perform competently in a profession from those who cannot. This is the ultimate priority of both the Examination Board and Professional Testing Services that provide standardized licensing examinations. According to Michael T. Kane, the kind of abilities tested should be considered critical abilities for a profession. He maintained that:

--Abilities should be clearly related to client outcomes.
--Abilities should be weighted according to their importance for practice.
--The level of proficiency required by the examination should not be higher than that required for practice.


5Ibid., p. 9.
—Extraneous factors should not unduly affect examination scores.
—Abilities tested should fall within the profession’s scope of practice.  

The faith of the public in testing has been reinforced by the recent tremendous growth in the use of tests in many sectors of American life. Despite critical articles and books that appear from time to time, many institutions in our society find test results useful in decision making.

According to Michael T. Kane, there are at least two possible interpretations for the scores on licensure examinations:

First, they can be interpreted as providing predictions of an examinee’s future professional performance. Second, the scores can be interpreted as providing evidence of an examinee’s present competence on specific abilities that are considered critical for practice.

However, the writer also believes that other factors such as: test validity, test reliability and test bias may also affect a candidate's performance on a licensing examination. Therefore, the two possible interpretations of a score on a licensure examination may not always be indicative of one's abilities. Thus, the criteria that are used for judging or assessing an individual's present or future performance require critical yet sensitive evaluation.

Since the early 1980s licensure examinations and standards have warranted special attention. The quality of licensing examinations and their objectivity have been taken for granted, especially when testing is done under public auspices. According to Susan E. Holmes,

---

7Ibid., p. 914.
credentialing examinations warrant special attention primarily for three reasons. These are:

1. First, the outcomes of credentialing examinations have an enormous impact on the lives of individuals, and on society in general. Such tests have associated with them the potential threat of serious harm to the public as well as the individuals who have invested a considerable amount of time, money and effort preparing for a particular profession or occupation.

2. Second, the number of professions and occupations currently requiring some kind of credentialing is greater than ever before, and the figures continue to increase.

3. Board members usually hold the major responsibility for these examinations. They are often individuals who do not possess the training and experience in test construction or the statistical analysis of test items.8

Currently, one of the major problems that has been associated with testing and licensure is that of test bias. A test is biased for a given group if it does not permit individuals the opportunity to demonstrate their skills as completely as it permits other groups.

Joan L. Herman, states that:

Bias can be apparent in a test in a number of ways, including obvious presentation defects (e.g. items that disparage some groups, or those that depict majority customs or activities that are stereo-typed etc.), linguistic and semantic problems and socio-cultural and contextual bias.9

A careful item analysis can minimize the more obvious problems, but such analysis should be supplemented with statistical procedures for


detecting bias. In addition, Herman introduces four inter-related qualities that are essential for developing reliable examinations for licensure. These are:

1. **Content validity**: Do the tests measure meaningful and significant competencies?

2. **Technical quality**: Are the test items technically sound, reliable, sensitive to instruction, and free from bias?

3. **Standard setting procedures**: Were reasonable procedures used to establish minimum performance criteria? Are the cut-off scores defensible?

4. **Curricular validity**: To what extent are the test competencies reflected in the instructional program?\(^\text{10}\)

As Dr. Shimberg points out, in Hetch's article, "Overview of Problems Involved in Validating Professional Licensing and Certification Examinations."

Despite a growing concern for licensure and validation in particular, there is a surprising lack of information and research on the topic. This is especially true in attempting to relate licensure to job performance.\(^\text{11}\)

In examining the role and functions of testing, it is useful to define the three direct participants in the testing process. They are the test producer or developer; the test user, usually an institution that expects to base decisions at least in part on test results; and the test taker, the individual for whom the test establishes a particular performance score. The producer develops tests that sample

\(^{10}\)Ibid., p.5.

performance, typically with a view either to establish a standard of a desired level of competence or to predict later performance in school or on the job. The test user is typically an educational institution, employer or licensing agency. How these tests function may vary with the various types of institutions; however, the primary purpose of testing is to use the results as an objective measure of performance. For the test taker, scores derived on a test tend to focus on identifying specific skills and abilities.

Testing is a very serious reality in our society. Testing works best for society and for everyone when it is used in conjunction with human wisdom, experience and knowledge. It is at its worst when relied upon too much, placing human discretion, judgement, and consideration in the background. Critics maintain that future performance can be predicted only in the most general terms, and to ignore this limitation by evaluating licensure examinations in terms of their ability to provide predictions of future performance is to set up an unattainable standard.
IV. METHODOLOGY

The methodological approach utilized in conducting this study is inductive analysis. According to Michael Q. Patton,

Inductive analysis means that the patterns, themes, and categories of analysis come from the data; they emerge out of the data rather than being imposed on them prior to data collection and analysis.12

In addition, the inductive analysis assisted the writer in describing many of the problems associated with testing and as a direct result, describing many of the problems that are commonly associated with licensure examinations.

As an intern with the State of Georgia's Examining Boards Division, the writer had the opportunity to directly observe the process involved in test development. Thus, the writer has chosen the Test Development and Test Administration Section as a frame of reference. The primary rationale for studying this problem during this time is that society, businesses, and governmental institutions are beginning to rely more heavily on testing for predicting future success, potential, promotions and employment. However, as the reliance on testing increases, exactly what these examinations can predict is of concern for many who are required to take these examinations.

A. Primary Data Collection Techniques

The writer utilized participant observation and informal conversational interviews as the main sources of primary data for this study. "A participant observer is one who joins a group or organization and acts as a typical participant in order to study the behavior of the group or organization." The writer interviewed Dr. Myrtice Wills-Dye (Executive Director), Mr. Grady Barnhill III (Senior Test Consultant), and Mrs. Celeste Gosier (Test Consultant). All three individuals work in the Test Development and Test Administration Section of the State of Georgia Examining Boards Division. These individuals were selected for the interviews based on their extensive knowledge and familiarity with developing licensure examinations. In addition, these individuals are experts in the field of test development, item construction and item analysis. Like participant observation methods, informal conversational interviews are used in an effort to obtain a more intensive study of perceptions, attitudes and motivations than a standardized questionnaire permits. This type of interview is useful in scouting a new area of research, to find out what the basic issues are and how people conceptualize the topic.

B. Secondary Data Collection Techniques

Secondary data were collected from a variety of sources including books, articles, lectures, and journals.

---

V. ANALYSIS

As stated earlier in this study, the writer has identified four areas of testing that impact greatly on the credentialing process of occupational licensure. Thus, the writer's analysis involves the discussion of these areas which are test validity, test reliability, test bias, and test policies. The writer has further provided a conceptual overview of these four essential areas within the discussion for the purposes of providing a background for comparing and constrasting the theory and application of these concepts.

The writer employs the use of inductive analysis in an effort to describe some of the issues that currently confront testing as an instrument for measuring minimum competency as it relates to occupational licensure.

For many people, the idea of evaluating individuals by tests is offensive. Attempts to numerically represent aspects of human behavior are seen as dehumanizing. Moreover, when contrasted to measurements in the physical sciences, testing is seen as subjective and impractical.14

Occupational licensure is largely the province of the states, although there is some local and federal involvement. Theoretically, states control occupational and professional licensing as an exercise of their

police powers, that is, for the protection of the public health, safety and welfare.

The quality of the tests for licensure and certification is difficult to ascertain, and also highly variable. Noonan and Nelson state that:

Tests and testing are the subject of intense controversy in American society. The signs of controversy range from polite disagreements among professionals about abstruse technical questions to heated public debates, with strong political overtones, about the social implications.15

The responsibility for determining the fitness of applicants for licensure is usually vested by law in a licensing board comprised entirely or predominantly of practitioners from the regulated occupation. At one time, most boards developed their own tests, and many of them still do. However, the most important function of a licensing examination is that it should reflect the level of knowledge, skills and abilities that an applicant is expected to possess in order to be credentialed.

According to Noonan and Nelson,

While not all tests are well developed and not all testing practices are wise and beneficial, available evidence clearly supports that proper use of well-constructed and validated tests generally provides a far better basis for making decisions about an individual's future performance.


16 Ibid., p. 15.
The broadest category of criticism contends that tests in general are neither sufficiently reliable nor sufficiently valid to justify their use. Some critics express dissatisfaction with the limited predictive powers of tests. Even when the validity of a test has been established that it measures adequately what it purports to measure - it is almost always validity for the short-term, and not validity for long term performance. In short, tests are seen by critics as being too limited in scope to measure complex characteristics of the kind required for long term prediction.

Another category of criticism focuses specifically on test construction, claiming that tests could be more valuable if only they were constructed properly. Critics point out that tests which incorporate a high verbal component, are used in nearly all testing simply because they are comparatively easy to construct and administer. Such use raises questions when the skill being tested does not require much verbal facility or fluency. Critics also argue that with multiple-choice items the "distractor," choices are often deliberately misleading or require overly subtle discrimination. In addition, there are also complaints that sometimes more than one multiple choice should be considered correct.

One of the most fervent criticism about test construction concerns the way in which test scores are "normed." It is argued that a test normed with members of the majority population will yield test scores that work to the disadvantage of test takers form other populations on whom the test was not normed. To overcome this problem, proponents of
tests and testing have long advocated, as a routine practice, the development of separate norms and the conduct of separate validity studies for majority and non-majority groups. Thus, the purpose of the test is to assess what the test taker will be able to do, after appropriate training, rather than to assess present knowledge.

All good measuring instruments have certain primary qualities in common. These qualities are validity and reliability. A test which lacks a known substantial degree of these primary qualities is not a measuring instrument in any true sense and little dependence can be placed upon results obtained by its use. In addition, there are certain secondary characteristics which are desirable in all testing situations: objectivity, administration, scoring and interpreting test results. However, these characteristics are considered far less crucial than validity and reliability, but they affect validity and reliability to some extent and in any event make the use of a test much simpler. Of the two primary requisites of good measurement - validity is generally regarded as the most important. The definition of validity in a testing situation may be illustrated by such questions as these: What does this test actually measure? To what extent does it measure a particular ability, quality or trait? In short, the question is essentially, to what degree does the test do the job it is intended to do? Ebel and Frisbie state that:

Validity traditionally has been regarded as a test characteristic, generally the most important quality of a test.... In particular, validity refers to the appropriateness of making specific inferences or of making certain decisions on the basis of scores from a test. The question, then is not 'Is this test valid?' but 'Is it valid to use the scores from
Methods for Estimating Validity

A variety of evidence might be presented to demonstrate valid test use, and most could be grouped into one of three categories: content validity, criterion-related validity and construct validity. The type of question to be answered by each is shown in table 1.

<table>
<thead>
<tr>
<th>TYPE</th>
<th>QUESTION TO BE ANSWERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content validity</td>
<td>How adequately does the test content sample the larger domain of situations it represents?</td>
</tr>
<tr>
<td>Criterion-related</td>
<td>How well does test performance predict future performance (predictive validity) or estimate present standing (concurrent validity) on some other valued measure called a criterion?</td>
</tr>
<tr>
<td>Construct validity</td>
<td>How well can test performance be explained in terms of psychological attributes?</td>
</tr>
</tbody>
</table>


According to Gerald H. Whitlock,

Current practice is to rely on content validity for professional examinations. There are marked differences among professions for measuring content validity and the quality of the procedures range from sloppy to excellent.18

---


Content validity is used to measure performance as evidence of performance in a larger domain of situations. A test is always a sample of the many questions that could be asked. Content validity is a matter of determining whether the sample is representative of the larger domain it is supposed to represent. Hence, the key element in content validity is the adequacy of the sampling.

There are two types of criterion-related validity. (1) Predictive validity, which is the use of test performance to predict future performance on some other valued measure called a criterion. (2) Concurrent validity which is the use of test performance on some criterion. The key element in both types of criterion-related validity is the degree of relationship between the two sets of measures (a) the test scores, and (b) the criterion to be predicted or estimated.

Construct validity is the use of an individual's test performance as a basis for inferring the possession of certain psychological traits or qualities. The aim in determining construct validity is to identify all the factors that influence test performance and to determine the degree of influence of each. The process includes the following steps: (1) identifying the constructs that might account for test performance, (2) formulating testable hypotheses from the theory surrounding each construct.

In discussing the two requisites of a good test, it is customary to think of validity as the most important quality. However, it is worth noting that reliability is essential to validity but the opposite is not so. A test may be reliable without being valid, whereas the validity of a test depends in part on its reliability; therefore the
validity of a test is limited by its reliability. Reliability refers to the consistency of measurement.

Unreliability, or inconsistency in a measuring instrument generally stems from two sources. These are, first, the situation in which it is used, including the physical and psychological state of the individuals tested, and second, the test itself. Such variable factors as conditions of testing, time limits and directions can be fairly controlled.

Methods for Estimating Reliability

The principal factors in the measuring instrument itself which may affect the reliability of a test are the quality of the individual questions or items and the length of the test. Concerning the individual items, there are many ways in which the quality of the questions can affect reliability. For example, a question may be ambiguous; that is, it may be subject to more than one interpretation or it may be so worded that its meaning is not clear.

The avoidance of ambiguity in items contributes materially to the attainment of reliability, though even the most skillful test makers cannot always avoid this fault. In preparing a test, one should guard against vagueness and eliminate those items which prove to be ambiguous. Practice and experience in making tests and a thorough knowledge of the subject matter are the best preventives against ambiguity in test items.

A second factor which is inherent in the test itself and which affects the consistency of measurement is the number of questions or the length of the test. Other things being equal the reliability of a test is proportional to its length; that is the longer a test is the more reliable it tends to be. Simply stated, the more samples that are
taken of a given area of knowledge or material, the more reliable one's assessment of that knowledge will be. The four basic methods of estimating reliability and the type of information each provides are shown in table 2.

**TABLE 2**

**METHODS OF ESTIMATING RELIABILITY**

<table>
<thead>
<tr>
<th>METHOD</th>
<th>TYPE OF INFORMATION PROVIDED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-retest method</td>
<td>The stability of test scores over some given period of time.</td>
</tr>
<tr>
<td>Equivalent-forms method</td>
<td>The consistency of the test scores over different forms of the test (that is, different samples of items).</td>
</tr>
<tr>
<td>Test-retest with equivalent forms</td>
<td>The consistency of test scores over both a time interval and different forms of the test.</td>
</tr>
<tr>
<td>Internal-consistency methods</td>
<td>The consistency of test scores over different parts of the test.</td>
</tr>
</tbody>
</table>


The test-retest method is essentially a measure of examinee reliability, and an indication of how consistently examinees perform on the set of tasks. The simplest and most obvious methods of obtaining repeated measures of the same ability for the same individuals is to give the same test twice. This would provide two scores for each individual tested. The correlation between the set of scores obtained on the first administration of the test and that obtained on the second administration yields a test-retest coefficient. The test-retest method is particularly useful in situations where the trait being
measured is expected to be stable over a period of time.

A number of objections to the test-retest method have been raised. One is that the same exact test items are used in both sets. Since this set of items represents only one sample from what is ordinarily a very large population of possible test items the scores on the retest provide no evidence on how much the scores might change if a different sample of questions were used.

Another objection is that examinee's answers to the second test are not independent of their answers to the first. Their responses on the retest undoubtedly are influenced to some degree by recall. Finally, re-administration of the same test simply to determine how reliable it is does not appeal to most licensing agencies.

Equivalent forms of a test are administered to the same group during the same testing session. If the multiple forms of the test have been produced in such a way that it seems likely that the scores on these alternate forms will be equivalent, and if each examinee in the group is given two forms of the test, then the correlation between scores on the two forms provides an estimate of their reliability. A high reliability estimate is evidence that the forms can be used interchangeably as measures of the same trait. The major drawback to this approach is that some testing agencies do not prepare alternate forms.

The test-retest with equivalent forms is a combination of the previous methods discussed. This is the most demanding estimate of reliability, since it takes into account all possible sources of variation. A high reliability coefficient obtained by this method would indicate that a test score represents not only present test performance but also
what test performance is likely to be at another time or on a different sample of equivalent items.

The internal-consistency method requires only a single administration of a test. A common approach is to split a test into two reasonably equivalent halves. These independent subsets are then used as a source of the two independent scores needed for reliability estimation. The standard method for splitting a test is to score the odd-numbered items and the even-numbered items separately. The correlation coefficient indicates the degree to which the two arbitrarily selected halves of the test provide the same results. Thus, this method is an indication of the internal consistency. Like the equivalent forms method, this procedure takes into account errors within the testing procedure and consistency over different samples of items, but omits the day-to-day stability of the examinee's responses.

Concept of Bias

In recent years, there has been a great deal of work with the construct frequently referred to as "Item Bias." Bias is a response to an item elicited by some trait on which two clearly identifiable groups differ markedly, but which is not related to the trait intended to be measured by the test. The issue of cultural bias arises from differences between the mean scores of different cultural groups on sets of items or tests. The rationale for invoking the concept bias rests on the assumption that the groups are actually equal in performance on the underlying construct being measured but that an unfortunate choice of items has caused the appearance of inequality in performance.

It is assumed by some critics of testing that there could be real
differences between cultural groups in test performance; however, caution must be exercised in defining bias. A difference between group averages does not automatically signify bias in testing. Thus, it appears that item bias is best defined as differences between groups in performance on an item that is greater than usual among items on the same unidimensional measure and this is due to irrelevant aspects of test items or the test itself.

Test Policies

Tests can be useful instruments, but some critics contend tests are clearly open to abuse - by producers as well as by the decision-making user, and even by the test takers. Hence, one of the most difficult policy questions to emerge from the controversy about testing concerns regulation: If tests continue to be used, what can society do to prevent their misuse? According to John Weiss:

Public policy protects the food we eat, the products we buy, the safety standards at our workplaces. Yet despite its overwhelming impact on both citizens and institutions, America's multi-billion dollar testing industry is virtually unregulated. Few consumer protection laws apply, and test publishers, refuse to voluntarily produce elected government officials, independent researchers, or test-takers with information needed to verify that their exams are fair and valid.19

In almost any profession or industry, some form of control or practices exists to maintain standards and to prevent abuse. Some observers have suggested that the testing industry itself should take a far more active role in combating test abuse. For producers, self policing

might include providing instructions in test administration and interpretation and allowing independent researchers greater access to test data, particularly validation data.

Interpretation of Analysis

The most fundamental challenge of the State Examining Boards Division for occupational and professional licensure examinations is validity. The agency is legally charged with establishing a testing standard that is acceptably valid for the purpose of ensuring functional competence. Only if this is achieved will the standard provide a measure of protection to consumers of licensee services. In the words of Grady Barnhill, Senior Test Consultant:

Although our responsibility is publicly understood to be competence evaluation, our tests measure knowledge and ability. It is essential that the knowledges, and abilities reflected in our tests be derived from careful consideration of licensee work roles.20

This is quite a different approach to content validity from simply building tests that sample material from curricula of preparatory courses required for licensure. The often significant differences between legitimate goals of educational programs and those of licensing programs must be recognized. Testing standards should not be developed to validate the content of academic training programs. Licensing tests must go beyond curricula and instructional validity to assess readiness to perform important occupational services expected of licensee. Thus, the goal in test development is to create items that discriminate

20Interview with Grady Barnhill III, The Examination Development and Test Administration Section, Atlanta, Georgia, 27 July 1987.
between candidates who can perform competently in an entry level position and those who cannot. This is the ultimate priority of both the Board and professional testing services.

The greatest difficulty in item development is changing the focus of item writers from that of academic testing to one of job-relatedness. This problem is most often resolved through intensive training sessions for the test consultants and professional testing services. Each item written is reviewed by Board members to determine whether the item is job-related and reflect the most critical knowledge required for safe practice at the entry level. For tests developed at professional testing services, items are reviewed by test service editors for content validity, style consistency and possible bias. Items are further reviewed to ensure that no advantages are given to the test-wise candidates. Wigdor and Garner, state that:

There are some experts, however, who question the adequacy of even professionally developed occupational tests insofar as they measure skills more closely related to academic achievement and not job performance as the indicator of professional competence.21

Proponents of this point of view argue that the usual description of licensing tests as "job-specific" or "job-related" achievements are misleading. In addition, critics also claim that licensure examinations are highly related to academic achievement, and these measures may not represent the minimum level of competency required for the practice.

According to Dr. Myrtice Wills-Dye, Executive Director:

With any testing situation one will encounter imperfect signs, because it is difficult to measure all knowledges, skills and abilities. In testing you are dealing with approximates of candidates' behavior. Hence, testing will work for some professions and not as well for others.22

Reliability measures provide an estimate of how much variation might be expected under different conditions. One general perception individuals have about test reliability is that there is a disparity between the passing rate on examinations that are administered more than once a year.

Grady Barnhill III, attributes the disparity and variability to what is termed "seasonal testing." Seasonal testing occurs when examinations are administered only a limited number of times during the year. For instance, the Certified Public Accountant examination is offered twice a year. The passing rate tends to be higher on the examination that is administered during the May testing session as opposed to the testing session in February. This is due primarily to the fact that individuals taking the examination in May tend to be first time test-takers. In addition, the May examination candidates consist of more recent graduates.

Some board-developed examinations lend themselves to questions of validity and reliability. This is due primarily to board members being forced to produce an examination, with little time in which to accomplish this goal, board members are forced to prepare test items

22Interview with Dr. Myrtice Will-Dye, The Examination Development and Test Administration Section, Atlanta, Georgia, 27 July 1987.
that may not necessarily accurately assess minimum levels of competencies which are required for licensure. However, for the most part board members attempt to prepare tests ahead of time to ward off future situations of this nature. Nevertheless, when this does occur the validity and reliability of the examination suffer.

Members of the Examination Section addressed the issue of bias in testing by indicating and supporting what relevant readings have suggested. No test can be completely free of bias, and that bias in testing can only be minimized. In addition, the belief that tests lend themselves to bias is based on the understanding that the language of the dominant group may differ from the language of the sub-groups of the population. However, critics of testing contend that most test instruments reflect the language of the dominant group.

While the State of Georgia, Teacher Certification Test (TCT) is not an examination administered by the State Examining Boards Division, many of the controversies surrounding the validity and reliability of this examination for teacher certification impact on the issues that the writer raises about occupational licensure for the State Examining Boards Division, Test Development and Test Administration Section. The Georgia Association of Educators has challenged the use of the TCT on the basis that:

(1) Teachers should not have to pass a test after they have been certified.

(2) A written test alone cannot evaluate a person's competence as a teacher.
(3) That a larger percentage of black teachers have failed the test than white teachers, thus, the test must be racially biased.\textsuperscript{23}

However, the State of Georgia has refuted the challenges and charges of GAE by stating that:

(1) Unlike other licensed professions, if an individual becomes dissatisfied with the practitioner they may opt to change to someone else. However, if a student is not satisfied with a teacher's performance he/she generally do not have the option to seek a different teacher.

(2) For teachers who fail the basic skills test, this is most likely a good indicator that someone is not a good teacher.

(3) Of 23,000 veteran teachers who have taken the test, only 327 have failed - less than 2 percent. Of the 327 that failed, 244 were black. However, more than 4,000 black teachers have passed the test.\textsuperscript{24}

While the questions of predictive validity, reliability and bias have been raised by the GAE, the association has not been able to substantiate its claims. Therefore, the administration of the TCT will continue to be used to qualify the certification of teachers in the State of Georgia.

Standard setting is a highly technical process which requires the expertise of test specialists. The performance of a candidate is interpreted in terms of his/her relative standing within the group of candidates being tested. In an attempt to ensure a more consistent definition of the level of competence, a criterion-referenced approach

\textsuperscript{23}Paul Coverdel, "Testing is: Reasonable Way to Weed Out Incompetent Teachers." \textit{Atlanta Constitution}, 1 October 1987, Sec A, p. 19.

\textsuperscript{24}Ibid., p. 19.
to standard setting is utilized by the State of Georgia Examination Section. This approach involves interpreting candidate's performance on licensure examination in terms of a standard or "criterion" judged to represent the acceptable level of competence.

The writer observed during a board meeting what appeared to be an arbitrary setting of cut-off scores. During the interviews of the members of the Examination Section, the writer raised a question regarding testing policies in relation to setting cut-off scores. Board members indicated that while it may have appeared that the board established cut-off scores after reviewing the scores of the examinations, cut-off scores, in fact, are not arbitrarily set.

What was communicated to the writer is that each board has policies that permit them great latitude in establishing cut-off scores between 70-75 percent for passing. In addition, some boards exercise this flexibility more than others. However, the latitude involved in setting cut-off scores must fall within a range of the mandated 70-75 percent which would deem the pass rate valid.

To date, none of the licensing boards in the State of Georgia has ever been legally challenged in a court of law. However, this is not to say that the examinations administered have never been challenged. During the writer's internship, the writer observed that several different licensing board's examinations were challenged. While the writer is not permitted to name which licensing boards these were, she was granted permission to briefly mention the surrounding controversies of these examinations.

One examination in particular received a lot of attention because
there was a rather low passing percentage. Many of the candidates who had failed the examination challenged the construction of the test items. However, during the review of the examination and the correct responses, both Test Consultants were able to defend the construction of the test items. However, during the item analysis of this examination, it was discovered that perhaps too many questions with high levels of difficulty affected the candidates' performance on the examination.

While many of the issues that the writer has presented cannot be directly associated with the State of Georgia Examining Boards Division, these problems still remain unique to testing in general. Thus, it is reasonable to assume that many of the issues discussed are peculiar to many licensing boards. However, the members of the Examination Development and Test Administration Section did offer the following information in regards to their examination standards.

In an effort to achieve fairness and ensuring standards for minimum competency no candidate shall be denied an opportunity to demonstrate his/her competence in the examination process for any reason other than the competency-related criteria for education, experience and special training as set forth in law or rule. In addition, the environment during the examination shall be such that it provides no undue distractions and creates no unnecessary obstacles that might prevent the candidate from demonstrating those competencies required to meet the minimum standards for licensure or certification.
VI. CONCLUSION AND RECOMMENDATIONS

Standardized group testing is a product of mass society. It was developed because there was a need to assess the talents of large groups of people efficiently and at a low cost. The techniques that allow assessment in these conditions however, may impose constraints on the quality of the assessment itself. Large scale testing does not allow the flexibility of clinical testing. It cannot equal the advantages of long association in judging a person's abilities. Although a well developed test can be a reasonably good predictor of the performance of people in the aggregate, it may be a poor predictor of the performance of a given individual.

Another important problem involving all quantified information including test scores is that testing tends to dominate the decision making process for employment. Quantification encourages the dangerous illusion that what cannot be reduced to numbers can be left on the periphery of the decision making process. Test scores, like all other data, have limited dependability and significance, but are often used as if they were meaningful in every circumstance.

Testing is a very serious matter, particularly when its results can determine whether or not a person can make a living in his/her area of assumed competence. While it is difficult to assess a person's ability on an examination, testing has been and continues to be the dominant way of doing things.
The findings of this study raise many questions about assessing the minimum level of competency for occupational licensure and licensure testing. The writer is of the opinion that testing and evaluating levels of minimum competency warrant further research.

In light of the problems discussed in this study, the writer offers the following recommendations:

1. The State Examining Boards Division, Test Development and Test Administration Section must develop a more extensive and ready reference item bank.

2. Examinations that are known to lack substantial validity or reliability (e.g. short-term validity or reliability) should not be administered.

3. Examination Boards should be required by law to prove that their examinations are job-related and validation studies of these examinations should be accessible to the public.
APPENDIX A

ORGANIZATION CHARTS FOR THE STATE OF GEORGIA

THE OFFICE OF THE SECRETARY OF STATE AND

THE OFFICE OF THE JOINT SECRETARY
APPENDIX B

GLOSSARY OF TERMS
Glossary of Terms

Construct Validity - a measure of how well a test measures constructs or dimensions that are judged to be critical for job performance.

Content Validity - an estimate of how well the test mirrors or reflects elements of the job domain.

Criterion-related Validity - a validity strategy whereby a test or predictor is correlated with job performance or criterion.

Cut-off Scores - scores on a test or predictor below which a decision is made (not to hire) and above which another decision is made (to hire).

Item Analysis - involves the counting of responses to objective test items to determine the difficulty and discriminating power of the items.

Norms - for a test indicate how the members of a particular reference group or groups scored on the test.

Predictive Validity - evidence for a test indicates how accurately some earlier measure of ability can forecast some later measure of performance.

Reliability - the consistency of a test or test items upon repeated measurement.

Tests - any paper-and-pencil, performance measure, or other information used as a basis for making an employment decision.

Test Bias - is a general term that represents a variety of factors or conditions that might give unfair advantage or disadvantage to individual examinees or groups of examinees. These influencing conditions may be present in the test itself, in the test administration, in the interpretation of scores, or in the use of the scores.

Test Items - is the smallest independent unit of a test. Each statement to be judged true or false, each question to which an answer is to be selected, each blank in a sentence or paragraph to be filled, each incomplete statement to which completion is to select a separate test item.

Validity - the degree to which a predictor or criterion measures what it purports to measure or demonstrating the job relatedness of a test by showing how well an applicant will perform based on the test predictions.
BIBLIOGRAPHY


Georgia, Official Code of Georgia Anotated, Title 43.


Interviews

Barnhill III, Grady. Senior Test Consultant of The State Examining Boards Division, Examination Development and Test Administration Section, Atlanta, Georgia. Interview, 27 July 1987.

Dye, Myrtice Wills. Executive Director of The State Examining Boards Division, Examination Development and Test Administration Section, Atlanta, Georgia. Interview, 27 July 1987.