DOCUMENT RESUME

ED 141 380 TM 006 110

AUTHOR Hecht, Kathryn A.; Fortune, Jim C.
TITLE Professional Licensing and Certification: Part I, Current Status and Methodological Problems of Validation. Part II, Analysis of Testing Procedures and Their Implementation for Practice.
PUB DATE 20 Apr 76
NOTE 35p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 19-20, 1976); For related document, see ED 110 491

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Certification; Criterion Referenced Tests; Evaluation Methods; Higher Education; *Occupational Tests; Performance Tests; Predictive Validity; Professional Associations; *Professional Occupations; *Professional Personnel; Professional Training; *Testing Problems; Testing Programs; *Test Validity

ABSTRACT

The need for better validation of professional licensing and certification examinations is discussed in Part I. Predictive criterion-related validation studies are difficult to develop, time consuming, impractical and expensive. Since training and knowledge variables are not necessarily sufficient to predict job proficiency, it appears logical to look to performance testing for solutions to the problems of validating and revising licensure exams more effectively. Four areas in which there are differences in practice across differing licensure examinations: (1) the type of test, (2) the type of controls involved in the testing process, (3) the kind of evidence utilized in the entry decision, and (4) the range of authority and the degree to which the examination provides a right to practice are discussed in Part II. Differences in practices across these four areas appear to be directed toward differential control of three primary sources of malpractice--absence of entry level professional skills, absence of adequate behavior patterns to cope with the pressures of professional practice and absence of adequate updating of professional practice. (MV)

PROFESSIONAL LICENSING AND CERTIFICATION:

(PART I)*

CURRENT STATUS AND METHODOLOGICAL PROBLEMS OF VALIDATION

Kathryn A. Hecht, Ed.D.
Center for Northern
Educational Research
University of Alaska
Fairbanks, Alaska 99701

Presented in the 1976 Annual Meeting of the
National Council on Measurement in Education
Symposium on Certification and Licensing in Competency-Based Education
April 20, 1976
San Francisco

*PART II: Analysis of Testing Procedures and Their Implementation for Practice, by Jim C. Fortune, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

As some of you may remember, Jim Fortune and I were involved in a symposium at the 1974 NCME meeting entitled Validation of Professional Licensing and Certification Examinations: A Methodological Dilemma. Jim was symposium chairman, and I delivered the overview paper entitled "Overview of Problems Involved in Validating Professional Licensing and Certification Examinations." I plan to use my time as both a review and follow-up of this initial paper, again highlighting some problems that I see in this area and looking for what changes might have occurred. I will cite examples from some of the same professions as discussed in the original paper and review what has happened over this two-year span. Jim will follow, citing additional professional examples and some additional methodological problems he has identified.

Copies of my original overview paper are now available through ERIC (ED110491), and since much of the information presented there will probably already have been covered by the time we get to this point in the symposium, I will only briefly review the points I thought were important both at that time and now.

In 1973, I became involved in the licensure validation issue when I was asked by the National League for Nursing to prepare a background paper on the validation of the RN (registered nurse) licensing examination and related work on performance testing. Naively, I thought it would be a simple task of consolidating what had been done in other professions. That was a far more complex and interesting task than I had expected, and questions and concerns raised during the assignment led me to urge Jim to organize a symposium on this topic. As far as I know, it was the first such discussion sponsored by NCME. I felt, at that time, that the symposium was quite successful,

and at least one other person who shared that platform with us, Ben Shimberg, is here today. (Unfortunately, Al Maslow, who was to be a discussant, was unable to attend - Al here's your second chance!) We had a highly attentive and interested audience although it was 5:30 before we were done. During the lively question-and-answer period, it appeared that several members of the audience were probably far more skilled and versed in this problem area than people on the panel. I was gratified by the amount of response the symposium generated; it was probably the most successful in which I have participated in ten years of attending AERA.

I hope this symposium and the audience contribution will be equally as stimulating. I am particularly pleased to see the issue arising again, two years after the first symposium, and consider it to be a follow-up to what we initiated, although the context seems a bit different.

Let me briefly tell you some of the things discussed in the first paper. First of all, I felt it was important to define the difference between licensure and certification, as I use those terms. I prefer Jensen's (1972) distinction where he discusses licensure and certification as two types of minimum competency testing in that the purpose of the test is to establish an individual's status in relation to an established go/no-go criteria. Licensing is usually a mandatory program designed to protect the public from incompetent practitioners; that is, to prevent an individual with particular deficiencies from entering practice. Jensen calls this a "selecting-out" process. Certification, on the other hand, is usually a voluntary program where the emphasis is on granting special status to an individual with more than run-of-the-mill knowledge, ability, and/or skill. Jensen calls this

"selecting in." Perhaps the best known example of a selecting-out exam would be for a driving license, where the public is protected from those whose knowlege is judged to be below adequate standards. An example of selecting-in, or certification, would be the "diplomate" program for medical specialities or recent certification programs for automobile mechanics. Since validation deals with the purpose to which the test is intended, I believe these distinctions to be important. Unfortunately, the words are often used interchangeably. For example, I belive teaching certification to be a misnomer, according to this definition, because it is a legal requirement to begin teaching, to protect the public from incompetent teachers, and signifies no special standing within the profession.

Next, I took a brief look at the sudden growing interest about validation of licensing and identified four concerns. First was the criticism of testing in general, which in the past decade, has become a popular, head-line making cause. Second, there was a proliferation of jobs requiring licensing and the hodgepodge of state and local legislative bodies emerging to control the process. (Ben Shimberg's 1972 report entitled Occupational Licening and Public Policy, is the only comprehensive document I have seen on the subject and it provides an excellent overview.) Third, the civil rights movement continued to make inroads against discrimination, especially here concerned with discriminatory practices in hiring. Fourth, challenges were being made to many professions to obtain status through alternative training routes rather than the traditional curriculum or school-based methods. This is a question of who qualifies to take a licensing exam. In 1972, for instance, many returning army medics sought to take the RN exam and were

denied such access on the grounds of not having graduated from nursing school. I think it is still safe to say that licensing is experiencing a period of attention and questioning for a number of reasons, not the least of which is federal legality. Licensing agencies continue to be subject to challenge to prove their tests are valid predictors of job performance, measuring job-related skills.

Despite the growing concern about the topic, I found an incredible lack of information, especially research information. This was especially true in trying to relate licensure to job performance. The information I was able to locate was scanty, often considered in progress, and being done in subject matter areas rather than considered collectively as a methodological problem. Very seldom was material available through professional journals and, in some cases, professions considered such information confidential.

Let me share with you something that Maslow, one of our discussants, said in 1971 at the Civil Service Commission Research Center, when he was advising the Council on Occupational Licensing (p. 339):

> *I am convinced that we need to sharpen our ability to develop and demonstrate the rational relationship between the job requirements and the measurement system used to certify or qualify people for an occupation. A number of techniques are available to improve the process of job analysis to get a much more exact fix on the critical requirements for the work to be done. I would urge, therefore, that especially in examinations for occupational knowledge and proficiency, you insist, at the very least, on a clearcut showing of how one proceeds from the decision as to the skills and abilities required for effective performance to the decisions that certain or other measures will insure that the applicant can adequately perform in that occupation.*

Let's see what changes there have been on some of these points over the past two years. First of all, there has been some professional attention paid to the matter which I had found extremely lacking previously. Let

me cite the APA publication Principles for the Validation and Use of Personnel Selection Procedures (1975) as an example. Issued by the Division of Industrial-Organizational Psychology of APA, it is intended to accompany the APA publication Standards for Educational and Psychological Tests (1974). However, in complimenting this body on issuing the Principles, let me also raise a point. This publication says it intends to provide: (1) principles upon which personnel research may be based; (2) guidance for practitioners conducting validation; and (3) information which may be interpreted for personnel managers and others who may be responsible for authorizing or implementing validation efforts. However, it also says that the Principles are not intended to be a technical translation of existing or anticipated legislation. I note this publication as evidence of heightened professional interest; but still there seems to be a reluctance to relate methodological principles, professional outlook, etc., to the predicaments in which many professions now find themselves in relation to validating licensing and certification exams. We still seem to have methodologists in one corner and applied researchers or practitioners in the other. Practitioners are still working in subject fields, and, most methodologists are still dealing with general principles. I have found little evidence that the "Twain has met" in the past two years.

Let me comment also on the availability of information. It appears to me that there is as little overall information available now as there was two years ago on how one should go about validating a licensing or certification exam, and on methodological issues related to these procedures. In two years, I have seen little in the professional journals dealing with this issue as a practical problem or even methodologically as a general

problem. Hopefully, other people during this symposium will have identified work I have been unable to locate.

Another thing that doesn't seem to have changed is the trend toward the proliferation of licensing exams. I admit that most of these are not what one generally calls professional; for instance, I went to the Reader's Guide (1973-74) and I found, under licensure, articles with the following titles: "Should Auto Mechanics Be Licensed by Law? Pro and Con Discussion;" "Licensing Proposal for Parenthood;" "Debate Over Licensing for Consultants;" "How to Strike Back Against Appliance Repair Abuse -- Suggestion for Mandatory Licensing;" "Drivers License for Snowmobiles?" So the problem of building and validating licensing tests doesn't seem to be going away--it seems to be increasing!

On the professional side, there appears to be some evidence of growing interest in certification exams--exams that grant recognition to a specialty area. For example, the American Nurses' Association is now in the process of developing a series of certification exams in nursing specialities, such as one already available in Psychiatric-Mental Health Nursing (1974). In 1971, National Association of Social Workers introduced a written objective examination for certification in their field. I'm sure there were numerous others.

In terms of legal challenge, I really don't feel adequately versed to deal with this issue, and by this point, hope it has been covered by somebody else on the panel. However, the work of the Equal Employment Opportunity Commission, the Civil Service Commission and others involved is very important to our topic and needs to be carefully observed and interpreted. Unfortunately, information on what is happening with regard

to the legal situation--and especially how that might relate to methodologists--is also scarce.

The Civil Rights Digest, a useful and free publication, devoted its Spring, 1975, issue to Job Discrimination and Affirmative Action. It includes an article by White on "Testing and Equal Opportunity" which notes that (p. 49):

> *In its recent decision in Moody v. Albemarle Paper Co., the Supreme Court kept up the momentum of equal employment opportunity by favoring compensatory back pay awards and by specifying "appropriate standards of proof" in determining whether tests are job related.*

The author was actually referring to a 1973 circuit court opinion on Moody v. Albemarle which said, with respect to the testing question (p. 139):

> *We think Albemarle has failed in several respects to show that its tests are job related, have a manifest relationship to employment, and have been validated in accordance with EEOC guidelines. . .*
>
> *In developing criteria of job performance by which to ascertain the validity of its tests, Albemarle failed to engage in any job analysis. Instead, test results were compared with possible subjective ratings of supervisors who were given a vague standard by which to judge job performance. Other courts have expressed skepticism about the value of such ill-defined supervisor appraisals.*

In June, 1975, the Supreme Court upheld and expanded upon circuit court ruling that Albemarle had not proved the job relatedness of its testing program: This decision gave great weight to the EEOC Guidelines (1974) and also referenced the APA Standards (1974). The court quoted the Guidelines concerning use of rankings (p. 25):

The Guidelines provide (that) the work behaviors or other criteria of employee adequacy which the test is intended to predict or identify must be fully described; and, additionally, in the case of rating techniques. . .whatever criteria are used they must represent major or critical work behavior as repeated by careful job analysis.

Let me now address the problem as I see it and as I view it over the two year period. First of all, validation studies of licensure exams are rare indeed. Seldom is the test development process that sophisticated or comprehensive. Often what is offered is some sort of content validation methodology, which varies as widely as calling in a group of professional cohorts to review questions to doing fairly adequate analysis of critical skills. Seldom are such exercises reported except to say they exist.

However, it would appear to me that predictive criterion-related validation studies would be the type most closely fitting the expressed purpose of licensure exams, that of assuring minimal competency on the job for the protection of the public. Interest is with the criterion not yet obtainable at the time of testing, as one wishes to predict an individual's outcome prior to a situation occurring--that being when the person is on the job. Of course, it can be said that for licensure exams, such validation studies, the predictive type, are difficult to develop, time consuming, impractical for numerous reasons, and expensive. It can also be said that psychometric methodology offers little guidance for such studies; the area of licensure lacks classic studies familiar to those who are schooled in psychological testing. Once all this is comprehended, the fact that such validations are rare, almost non-existant, is not surprising but nevertheless disconcerting.

Research has shown that course grades are not efficient predictors of job performance (best demonstrated by the Navy in World War II: Stuit, 1947; Gulliksen; 1950). We know, for instance, that many written tests are found to correlate only too well with reading tests in non-professional skill areas. We know there is a lack of relationship between achievement as measured by paper and pencil tests and performance measures. This has been demonstrated in such diverse professions as education (Quirk and others, 1972) and engineering (Hemphill, 1963). This lack of correlation suggests great importance for validation of licensure tests. As far back as 1951, Ryans and Fredericksen clearly summed this point up from the measurement perspective (p. 455):

> *From the standpoint of validity one of the most serious errors committed in the field of human measurement has been that which assumes the high correlation of knowledge of facts and principles on the one hand and performance on the other. Nevertheless, examinations for admission to the bar, for medical practice, for teaching. . . are predominantly verbal tests of fact and principle in the respective fields. Relatively little attention has been paid to the testing of performance as such.*

If training and knowledge variables are not necessarily sufficient to predict job proficiency, it appears logical to look next to what is broadly called performance testing. If one accepts the definition of performance testing as being a test which is "relatively realistic" (Fitzpatrick & Morrison, 1971), then this is a logical place to look for the answer to: (1) how to validate licensure exams more effectively, and (2) how to revise licensure exams themselves if necessary.

The most interesting and well documented use I found concerning performance measures in predictive research was in the area of employee selection and promotion. Of course, the employer situation has numerous

advantages over licensure boards, such as: control over subjects, limited range of jobs, job descriptive information, length of period of observation, and the possibility of gradually implementing a testing program, allowing research time to study predictions before changing the decision process. In this field, assessment centers are an effective performance-based type of employment or promotion screening device. The procedures stimulate situations that would be faced if the examinee were to be moved up to a higher grade position, and using a sophisticated form of role playing, develops information regarding how he might cope with decison making. Validation studies done on assessment center techniques have shown them to be a better indicator of future success than any other tool management has yet devised. (Bray and Campbell, 1968, describe how one such center works and includes the validation process.) Although the assessment center concept could be used as a validation tool for licensing exams, as an on-going technique, where large numbers of people must be tested each year, it is obviously unrealistic.

Illustrative of a more practical approach to introducing performance into testing situations are two types of programmed testing developed in the medical profession. In the assessment of clinical competence, performance is simulated on objectively scored paper and pencil tests. The National Board of Medical Examiners first introduced the concept (Hubbard, 1964) and now use programmed testing for the medical licensing exam Part III on clinical competence, which previously was a practical bed-side type of oral examination. There are two competing models in the medical profession for this type of testing. In both models, the examinee is confronted by a realistic clinical situation and proceeds through a series of

decision choices, each step accompanied by an increment of information upon which the next depends, similar to programmed teaching. The linear model used by the National Board allows for only one correct option at each decision point while the branching model promoted by McGuire and Babbott (1967) allows for more than one route to a solution and is used in specialty certification.

To my knowledge, neither variation has been validated in relation to predicting job performance. The Part III, or clinical competency exam, is said to derive its validity by measuring something different from what Parts I and II measure, which is strongly related to medical school course work; scores on Part III correlate only moderately with those on Parts I and II. Cronbach (1970, p. 444), having reviewed this so-called validity evidence on Part III notes: "Follow-up studies are needed to make sure the test measures a skill of medical practice and not just ingenuity in test taking."

Similar to problems confronted by those attempting predictive validation of licensure tests, performance tests development logically begins with an identification of specific skills and abilities involved in the activities the test is designed to predict. The next step is the choice of representative tasks, a difficult task which strongly influences the validity of the performance tests. Other difficulties with performance testing come from a lack of applied methodology in that performance tests are by nature criterion-referenced and procedures for estimating reliability and validity are still subject to debate.

I would like to digress on this last point, a problem area which I

has been seriously neglected by methodologists. As I noted previously, licensure and certification exams are types of minimal competency exams and like the performance measures we have been discussing, would normally be considered criterion-referenced. The examinee is theoretically tested in terms of absolute criteria; how well the person can do what he or she needs to do to perform adequately on the job. Comparison among test takers is not the purpose of licensing exams. However, most licensing exams on which I have reviewed research data are still developed on norm-referenced models.

Returning to performance tests, most of the literature discusses them as a new form of assessment, used to increase the realism of the test. My primary interest in performance tests is less commonly discussed, that of providing criteria for predictive validation. The only suggestion of such research I have located was a theoretical discussion on "Providing a Criterion Measure" also by Ryans and Frederiksen (1951, p. 466):

> *Performance test data may provide, first of all, a criterion for research. Information yielded by performance tests makes possible the validation of other measures which, although of a more indirect nature, may be more economical in administration. In many situations, it is difficult and expensive to administer performance tests to large numbers of examinees. Such situations demand the construction of psychometric instruments that will yield measurements related to criterion and will be also practicable. In the construction of aptitude tests for various skills and operations, performance tests may provide the criterion against which the available second order test can be judged.*

Looking to specific professions and validation efforts, I would first like to thank the National Board of Medical Examiners, Educational Testing Service, and the National League of Nursing for supplying me with the information for this section. The information was obtained through personal contacts and most is not generally available.

The National Board of Medical Examiners appears to have no new validation studies available on the National Board Examinations Parts I, II, and III for medical students. (The two studies sent to me were both over ten years old.) However, in 1974, the National Board formed a R & D Department, and their correspondence says several developmental studies which include validation components are in progress.

Two validation studies of certifying exams in the medical profession I can report on concern practitioners of internal medicine and physician assistants. Schumacher (1973) reports a valuation study of the written examination for internal medicine which was based upon comparing test performance of first and third year residents. According to the author, the study attempted to (p. 132) ". . . determine whether the examination is sensitive to gains in knowledge, ability, and skill that presumably occur during graduate training in internal medicine." The study was said to (p. 133) ". . . support the hypothesis that the examination had validity for measuring achievement at the graduate level." The study makes no pretense of relating achievement to expected future performance or even a sound job analysis.

The Physician Assistant Certifying Exam, new in 1973, is also said to have validity in that those with formal training scored higher than those just beginning training. However, the report (National Board of Medical Examiners, 1973-74) also noted evidence of construct validity in that postgraduates with clinical experience after graduation did better on the exam than those without, though experience prior to entering training had not proven significant. An additional study to estimate concurrent validity compared program faculty rating of students on 40 statements of clinical competence with examinating. Four of the six examination components correlated significantly (though the highest was only .20) with at least one

of three rating factors. The report indicates modifications will be made in the following year to refine the rating scale. However, the report does not deal with predictive validity nor is there any mention of content or job analysis in developing the rating form.

The National League for Nursing[1] is proceeding with plans to validate the RN licensure exam, and in designing the project, has accepted ". . . as a first principle, [that] the sampling of content for a new test should be altered in such a way that items drawn from on-the-job experience would receive precedence over items based mainly on knowledge." The rationale given is that licensure is meant to assure safe and effective practice, and to evaluate nursing practice would require judgments as to whether the decisions and actions taken in sample cases is appropriate. In shifting from curriculum-based learning to job-related behavior, the question becomes how to sample items from the job situation. The League with A.I.R. (American Institutes for Research) is currently carrying out a critical incident study which will become the empirical base for a criterion-referenced test. Almost 14,000 incidents were collected in five specialty areas with geographic and institutional diversity. Classification of incidents and test specifications are the activities of the current phase.

The League then plans to construct a performance simulation test and compare results with the current RN exam. It is currently seeking funding to develop more elaborate performance simulation exams and to attempt validation by observation of performance. (This appears to be by the most active professional group on the validation scene to date!)

---

[1]This information is drawn from a draft proposal which the author has permission to use but which is not generally available and thus no reference is noted. (See Hecht, 1974, for additional information on the RN licensure exam and related studies.)

Looking to some professional exams that were not reviewed at the last session, Educational Testing Service has been involved in certification or licensure test development for social workers, lawyers, and nursing specialists.

In 1969, the National Association of Social Workers required the addition of an objective written test to the membership requirements of the Academy of Certified Social Workers. Since 1973, ETS has been administering that test. Research on the test reported by Boyd (1975) noted that no item clusters were large enough or differentiated enough to provide subscores--subjects doing well in one subject area would do well in another. Biographical data showed Blacks with consistently higher failure rates. Further study (Sharon, 1975) recommended reduction of jargon and wordiness, that study guides be provided, and that the test be empirically validated. A validation study for the social worker certification exam has just been approved. The information I was sent states the exam must grant certification to competent workers regardless of test-taking ability, and deny certification to incompetent workers regardless of their theoretical knowledge. (Sounds more like a delayed licensure exam!) The validation method outlined utilizes supervisor rankings on overall competence and on several specific dimensions of job performance related to a restructured exam. If pairs of raters agree, the criterion is assumed to be reliable and items thus developed to relate to effective job performance. No mention of job analysis is made. The plan also states, among utilization of study results, that such information will provide evidence regarding fairness in case of litigation.

A validity study on the multi-state bar exam is near completion (draft, Carlson and Werts, 1975) and will be reported on at both NCME and AERA meetings by Dr. Al Carlson. The purpose of the study was to investigate

the relationship among the Multi-state Bar, bar exam essays, law school grades, law school admission tests and undergraduate scores. The summary suggests reasonable relationships were found among these measures. Though this study does describe the relationships of the exam to previous achievement/aptitude measures of law students, it does not appear to be concerned with predicting adequate job performance. Therefore, the exam's validity as a licensing instrument appears in need of further study.

No research information was made available to me on the Psychiatric-Mental Health Nursing Certification Examination sponsored by the American Nurses' Association and developed by ETS.

Given the research information previously cited in this and the previous paper, it would appear that most professions have a long way to go in adequately validating licensure and certification exams. Few studies are directed at predictive validity or the performance which can be expected on the job. Many seem content to relate exams to academic measured based on curriculum content and training which have no proven relationship with the tasks to be performed. Most exams are put into use before validity studies are started or with none in mind until problems identify themselves. And those professions I have cited here are probably further along than most. But given recent legal decisions, time seems to be running out.

An additional challenge to professional individual licensure is institutional licensure. Interestingly, it appears most vocal in the one profession which has appeared to take leadership in licensure validation--nursing and other health care occupations. Agree (1973) reviews the movement in health care and notes support from the Department of Health, Education, and Welfare in testing various plans of institutional licensure

for health personnel. Both the A.M.A. and the N.L.N. have formally opposed institutional licensure in lieu of individual licensure for doctors and nurses. If nothing else, this alternative should further prompt action toward upgrading licensing and certification programs and can be viewed as an added critical pressure on professions to police themselves wisely and equitably.

In closing, I think the field of licensing and certification continues to present methodologists with real and immediate challenges. Here are practical problems based on real and current concerns. If each occupation continues struggling on its own, without serious attempts from a group (such as we have here today) to provide integrated conceptual and methodological frameworks, solutions will remain a long way off.

In two years, since I last wrote that closing statement, I have seen little to indicate a cohesive methodological effort being made to tackle any of the problems mentioned herein. Professional licensing and certification exams continue to be developed and used without adequate validation data. What work that is being done tends to be the dilemma faced by those responsible for licensure and certification. Few share their experiences or the knowledge gained. Hopefully, panels like this will create more interest in this area of applied methodology. If there is another panel in two years, we hope to see significant progress being made.

BIBLIOGRAPHY

Agree, B. C., "The Threat of Institutional Licensure," American Journal of Nursing, October, 1973, Vol. 73:10, pp. 1758-1763.

Albemarle Paper Co., et al. v. Moody, et al., ___US___, 45 L Ed 2nd 280, 95 S Ct 2362 (June, 1975).

American Nurses' Association, "Certification Examination, Psychiatric-Mental Health Nursing--Bulletin of Information for Candidates, 1974," Princeton: Educational Testing Service, 1974.

Boyd, J. L., Jr., "Development of the Testing Program for the Academy of Certified Social Workers," Princeton: Educational Testing Service, September, 1975.

Bray, D. W., and Campbell, R. J., "Selection of Salesmen by Means of an Assessment Center," Journal of Applied Psychology, 1968, Vol. 52:1, pp. 36-41.

Carlson, A. B., Werts, C. E., "Relationships Among Law School Predictors, Law School Performance and Bar Examination Results," Princeton: Educational Testing Service, Draft--October, 1975.

Cronbach, L. J., Essentials of Psychological Testing, (Third Edition) New York: Harper and Row, 1970.

Dowaliby, F. J. and Andrew, J., "Relationships Between Clinical Competence Ratings and Examination Performance," Philadelphia: National Board of Medical Examiners, (no date).

Fitzpatrick, R., and Morrison, E. J., "Performance and Product Evaluation," In Thorndike, R. L. (Ed.), Educational Measurement (2nd Edition), Washington, DC: American Council on Education, 1971.

Guliksen, H., "Intrinsic Validity," American Psychologist, 1950, Vol. 5, pp. 511-17.

Hecht, K. A., "Validation of Professional Licensing and Certification Examinations: A Methodological Dilemma," paper presented at the Annual Meeting of the National Council on the Measurement in Education, 1974, ED110491.

Hemphill, J. K., A Summary of the Engineering Study, Princeton: Educational Testing Service, 1963.

Hubbard, J. P., "Programmed Testing in the Examinations of the National Board of Medical Examiners," in 1963 Invitational Conference on Testing Problems, Princeton: Educational Testing Service, 1964.

Jensen, O., "Some Thoughts on Minimum Competency Interpretations of Performance on (M-C Item) Achievement Tests," (Mimeo), Educational Testing Service, Princeton: 1972.

Maslow, A. P., "Licensing Tests-Occupational Bridge or Barrier?," Proceedings Third Annual Conference, Council on Occupational Licensing, July, 1971, (see Shimberg, 1972).

McGuire, C. H., and Babbott, D., "Simulation Techniques in the Measurement of Problem-Solving Skills," Journal of Educational Measurement, Spring, 1967, Vol. 4:1.

Moody, et al. v. Albemarle Paper Co., et al., 474 F. 2d 134 (4th Cir. 1973).

National Association of Social Workers, Inc., "Academy of Certified Social Workers--Information Bulletin, 1975," Princeton: National Association of Social Workers, Inc. and Educational Testing Service, 1975.

National Board of Medical Examiners, "National Examination Program for Assistants to the Primary Care Physician--Project Report, 1973-74," Philadelphia: National Board of Medical Examiners.

Principles for the Validation and Use of Personnel Selection Procedures, Dayton: The Industrial-Organizational Psychologist, 1975.

Quirk, T. J., Witten, B. J., and Weinberg, S. F. A Critical Review of Research Related to the National Teacher Examinations, Princeton: Educational Testing Service, 1972.

Ryans, D. G., and Frederiksen, N. "Performance Tests of Educational Achievement," In Linquist, E. F. (Ed.), Educational Measurement, Washington, D. C.: American Council of Education, 1951.

Schumacher, C. F., "Validation of the American Board of Internal Medicine Written Examination," Annals of Internal Medicine, January, 1973, Vol. 58:1, pp. 131-135.

Sharon, A. T., "Assessment of Competence in Social Work: A Report on the 1973 Assessment Results of the Academy of Certified Social Workers (ACSW) Program," Princeton: Educational Testing Service, 1972.

Shimberg, B., Esser, B. F., and Kruger, D. H., Occupational Licensing and Public Policy, Princeton: Educational Testing Service, 1972.

Standards for Educational and Psychological Tests, Washington, DC: American Psychological Association, 1974.

Stuit, D. B. (Ed.) Personnel Research and Test Development in the U.S. Bureau of Naval Personnel, Project N-105 and the College Entrance Examination Board, Princeton: Princeton University Press, 1947.

U.S. Government, "Guidelines for Employee Selection Procedures," Title 29 Labor, Chapter XIV, Equal Employment Opportunity Commission, Federal Register, July, 1975, Part 1607, pp. 807-12.

PROFESSIONAL LICENSING AND CERTIFICATION

PART II

ANALYSIS OF TESTING PROCEDURES AND THEIR IMPLICATIONS FOR PRACTICE

Jimmie C. Fortune, Ph.D.
College of Education
Virginia Polytechnic Institute
and State University
Blacksburg, Virginia 24061

Presented at the 1976 Annual Meeting of the
National Council on Measurement in Education
Symposium on Certification and Licensing in Competency-Based Education
April 20, 1976
San Francisco, California

In the second part of our paper, we present an analysis of the licensing certification processes as they are practiced by the different professions. This analysis includes an investigation of different procedures used in the development and the administration of the licensing examination. The investigation will focus upon the roles and benefits of particular practices and will attempt to determine their implications for the educational professions and for competency-based education.

An Analysis of Licensing and/or Certification Examination Processes as Practiced by Different Professions

I would like to look at the variety of practices which exist in four different areas of the licensing examination process. These four areas are: (1) the type of test; (2) the type of controls involved in the testing process; (3) the kinds of evidence utilized in the entry decision; and (4) the range of authority and the role of the examination as it relates to right-to-practice.

Type of Test. Four factors should be considered in describing the differences in the type of test or licensing examination used by different professions. These four factors include: (1) the construction of the test; (2) the measurement strategy used in the test; (3) the type of setting in which the test is given; and (4) the scoring process. With regard to the construction of the test, several practices are used by various professions. In some professions, the licensing or certification examination is made up by trainers in that profession. Such examinations usually reflect the training curriculum which has been required for entry into the given profession. The content of other examinations is provided by practitioners

already in the profession. This kind of examination is designed to test on content which comes from practice of the profession and which is deemed necessary for successful practice in that profession. A third option in the construction of professional licensing examinations is the development of the test through contract with a professional test development firm or psychometrician. The developers of the test usually use content extracted from the training curriculum and from work samples taken from the field of practice. Still a fourth option in the construction of the test is to take items from either a theoretically-based or empirically-generated set of professional competencies. In some cases, these competencies have come from panels of experts--both practitioners and trainers--who were asked to identify competencies through some systematic procedure of brainstorming. In other cases, competencies are derived from task analyses of professional records or behaviors.

Each of these four processes of test development maintains a certain degree of face validity relative to the measurment of skills considered by an identified reference group to be essential for professional practice. These different routes tend to generate differences in content but appear to offer equal opportunity to utilize appropriate test construction techniques. The various examinations, however, do not appear to have undergone the same level of test analysis or to have been developed at the same level of psychometric sophistication. Present legal issues indicate a need to gather more evidence on the validity of the examinations for individual assessment uses.

The second consideration in regard to type of test is the measurement strategy espoused by the examination. Most of the licensing examinations used by the professions are fixed-response, machine-scorable, norm-referenced

examinations. This form of examination has been adopted primarily for convenience. Also, there is a need in licensing examinations to maintain security of the test. The most convenient way to maximize security is to use a random, evolving-item pool; and multiple-choice items are the easiest to construct for item-pool utilization. Moreover, in licensing examinations, there is a need to process a large number of tests rapidly and accurately. Speed and accuracy are assured by machine scoring. The norm-referenced characteristics of these exams probably came into being because the infancy of criterion-referenced measurement and logistical difficulties of performance-based tests limited the use of these alternatives. It is anticipated that test strategy options based upon performance and criterion-referenced testing will become practices of the future since the measurement assumptions in such strategies are so appropriate to licensing examinations. Given that the primary testing purpose is to guarantee the presence of given competencies, the criterion-referenced strategy of measurement is an extremely promising means to reform licensing examination procedures. Furthermore, it appears that increased evidneces of the predictive relationship between the licensing examination and competency in practice are essential to guarantee the civil rights of all. Performance testing offers one way to increase this predictive relationship.

The third factor to be considered in describing variations in type of test is the setting in which the test is administered. In some examinations areas where individual performance or simulated practice is required, the test is individually administered. Most often, however, the licensing examination is given in a public setting at a time of year selected for client convenience (e.g., the exam may be given immediately at the end of a training program or at the beginning of a seasonal practice period).

This kind of testing process involves travel of the examinee to the test site and the administration of the examination materials under standardized conditions by a professional examiner. Other examination settings might include the location of a particular training program or an employement center. The primary effects of the setting on the examination procedure appear to be related to the degree to which the examination environment is removed from the field of practice and to the degree to which the setting creates anxiety.

The fourth consideration to be made in terms of the type of test is related to the grading or scoring procedure. Some types of items requfre more subjective judgment for grading than do other items which are of the fixed response, factual content nature. Often, the items that require subjective judgment in their scoring can be hypothesized to contain greater content validity with regard to actual professional practice than do factual items. In the scoring of professional licensing examinations, practices range from the subjective judgment of peers, whose objectives are to protect their profession from entry by incompetents, to rote marking of mastery-type items dealing with skills or knowledge deemed essential for professional practice. Although the subjectivity-objectivity continuum of scoring is not directly associated with who grades the test, test scoring relative to who determines what is the right answer is another area in which practice varies by profession. As mentioned, the correctness of answers is sometimes determined by trainers, sometimes by practitioners in the field, and sometimes by mixed committees. Often the trainers foster entry into the profession; practitioners tend to policy entry more carefully.

The Type of Controls Involved in the Testing Process. A second area of practice that is worthy of concern in professional licensing procedures is contr over the examination process. The first consideration in looking at controls is the agency or type of group which is responsible for making the entry decision. Four kinds of groups typically police professions. Generally, the profession polices itself through an association. This association, such as the American Medical Association, frequently maintains control nationally over the licensing examination. In other cases, such as law, professions police themselves through state associations. Bar examinations, with the exception of subscribers to the multiple state bar examinations currently being offered by ETS, provide control of entry at the state level by colleagues in the law profession. Any advantage gained through local control may be sacrificed by the requirements of reciprocity and by the necessity to tolerate nonuniform standards of entry into the profession. A third type of control involved in the testing process is direct control by the state government, as is exercised by the certification of teachers. Although a professional association is indirectly involved in the certification process by recommending standards, certification requirements for teaching are established primarily by each state department of education. Often this certification right is delegated by the state department to training institutions who certify on the basis of successful completion of courses, rather than successful performance on an examination. In some circumstances, training institutions require a comprehensive exam- [illegible] potential graduates over a required curriculum. A fourth source of testing control is that which is available to a potential employer [illegible] the [illegible] examination [illegible] gained at a local level. [illegible] examin[illegible] is often given by employer's personnel offices and [illegible] battery [illegible] by a testing firm.

Two additional means of control have at times been practiced by professions. These two means of control, although used less often than the controls described above, do provide options which may, in some cases, be useful to the profession. The first of these options is control or certification through a subcommittee of a professional association. This subcommittee fulfills the certification or accreditation role with support from the profession as represented by association membership. The National Council for the Accreditation of Teacher Education provides a good example of this practice. The final element of control comes through a governmental agency in the form of a public protection law. Although not directly involved with a profession, licensing procedures for pilots utilize this type of professional entry control.

The essence of control of entry into a profession seems to lie either in the profession itself or in a subdivision of government. It would appear that both state government and professional association controls appear useful and essential in cases where public safety is an issue and where clientele may not have the choice of selecting the practitioners. Regional or state control usually results in non-uniform standards, which require the practitioners to undergo reexamination or to otherwise qualify for reciprocity before they can move their practice across a regional or state boundary. Hence, the uniformity of standards generated by national licensure control provides a mechanism through which professionals can avoid the nuisance of reexamination and/or states can avoid the risks involved with reciprocity. Most often, however, national licensure tends to occur only at the entry point in a profession, which fails to assure

up-date capability to keep practitioners in the profession aware of current developments in the profession.

A second consideration related to control over the testing procedures involves options available at the application level for professional examination. These options vary from required membership and petition for examination through a professional association to application and fee registration. Other options involve testing at the close of a training program or testing at the time of application for employment. In cases where fee application is required, more emphasis is usually placed upon the test results and there is more standard application of the test content to the entry decision. When the examinee must belong to a professional organization and must petition that organization for the right to be examined for entry into the profession, then it is more likely that prior qualifications and training become additional considerations for entry into the profession. In some cases, more weight is placed upon qualifications than upon the results of the examination.

The Kinds of Evidences Utilized in the Entry Decision. A third condition that should be examined in ascertaining right of entry into a profession is the type of supplementary evidence that is used in either determining eligibility for examination or for determining professional entry. Traditionally, training, experience, and course work have been major considerations in ascertaining a candidate's qualifications. In the case of law, completion of a trining program is required before the bar examination can be taken. In some cases, including teaching, the completion of specified course work with a minimal grade point qualification is all

that is needed to enter the profession. During the latter phases of the educational training program, potential teachers go through a period of practice teaching, which is usually one of the requirements for certification. This experience varies in the degree to which the practice experience approaches a longitudinal entry into the profession. In some cases, such as medical licensing, entry is possible only after the completion of both a training program and a term of supervised professional practice ("residence"). The residence in medicine more closely approximates professional practice. It can be hypothesized to provide a better gauge of how the potential practitioner will perform his or her career endeavor.

An additional source of evidence which may be examined prior to a candidate's entry into a profession is peer acceptance, which is usually gained through the solicited sponsorship of a candidate by a member already in the profession. The American Psychological Association, for example, requires sponsorship by two members prior to a candidate's admittance to that association. A candidate usually earns peer acceptance through his association with members of a profession during this training or apprenticeship.

The Range of Authority and the Role of the Examination as it Relates to Right-to-Practice. It appears that in licensing examinations the range of authority varies from simple certification or accreditation by the profession to actual blockage of practice in the field. For instance, non-credentialed teachers may teach with only a minor penalty in terms of compensation for services rendered. Licensing or certification of psychological testers is based upon a graduated scale. Levels of examiner

sophistication are determined, and less qualified persons are barred from using some types of tests. In the medical profession, licensing is essential to any practice at a national level. Bar association tests tend to keep underqualified persons from the full-fledged practice of law.

It appears, therefore, that two types of power or authority can be associated with professional examinations. The first is associated primarily with the licensing examination, and is that of barring the unqualified from practice. The examination tends to establish the existence of a minimum level of skill and knowledge, which has been deemed essential for successful professional practice. The examination is intended to police the profession of unqualified people, and also to raise the probability of capable professional practice. Often, however, the examination ascertains evidence of minimal qualifications only and fails to produce predictive evidence of future success in practice.

A second power or authority assigned to certification examinations is the power to identify expertise in a speciality area. Such examinations are designed to provide evidences of special skills or knowledge that are currently held by practitioners in the profession. These examinations must be designed to demonstrate the mastery of special competencies or outstanding capabilities. In order to produce evidences of individual advancement, the examinations must possess the capability to discriminate among those professionals who have that special capability and those who do not.

Components of a Licensing Examination Process which are Functional in Licensing and Certification for the Professions

In order to identify what is required of a profession's licensing or certification process, it may be helpful to investigate the causes of malpractice by the professionals. Failure to provide adequate professional services can be traced to three principal causes. First, some professionals fail to stay current in their field, and the level of practice in their professions advances beyond their own level of expertise. Second, some professionals may enter the profession with an adequate set of functional competencies but may fail to make personal day-to-day adjustments to practice. In such a case, professionals may develop a personal problem or habit which interferes with professional practice and care for the client. And third, some professionals fail in practice because they lack adequate competence in one or more of the fundamental skill areas needed for successful practice in the profession. It is my opinion that this last reason for failure is reduced to a minimum by current examination practices.

It is believed that malpractice in our professions today stems primarily from one of the first two causes of failure. If this is true, what implications does this have for future practices in professional certification and licensing? What are the implications for the recent movement toward competency-based education and competency-based professional training programs? I think it becomes quite apparent that a periodic review or reexamination of professional practitioners can reduce the incidence of failure due to the first cause, the inability of the practitioner to stay current. Such rapidly developing fields as dentistry require

the practitioner to invest time and effort in keeping abreast with the field. Licensing at the close of a training program will not insure that professional growth and updating of practices will occur. Hence, it becomes apparent that reexamination of licensed professionals should occur with a frequency commensurate with the speed at which the profession is developing.

The second primary cause of failure in the professions, namely, the lack of adjustment to professional practice and the development of personal habits incompatible with good professional practice, cannot adequately be dealt with within the context of the licensing examination setting. Instead, this problem must be dealt with in the setting of professional practice. Here, the medical profession's model of using a period of residence seems to offer a potential solution. To translate this practice into the area of teaching, one might suggest that a teacher teach for three years under the supervision of an established colleague prior to his being given full professional certification. To some degree, it can be argued that the practice teaching experience satisfies this level of practice. However, the degree to which the practice teaching experience fails to simulate the actual teaching situation is directly associated with the degree to which practice teaching fails to produce the longitudinal personal evidence needed to predict success in practice. Residency requirements do provide a greater amount of longitudinal evidence than can be gained in an examination setting. Threats of disbarrment practiced by the legal profession can extend the guarantee of adjustment over a longer period of professional practice.

It appears that criterion-referenced testing procedures are also essential to complete the revision of the testing-out process. The criterion-referenced testing procedures allow for a more careful delineation

of competencies and for a direct association of test content in those competencies. Current practices in licensing fail to generate all of the validation evidence necessary to make accurate and unbiased assessments of individual competence and potential. The criterion-referenced testing strategy can, with proper validation, measure competencies which can be defended as fundamental to professional practice. It could be argued that the use of norm-referenced tests may be used primarily in the certification of exemplary proficiency and may in fact satisfy one aspect of the testing-in function of certification examinations. Such would not be the case, however, in the identification of mastery of specialized skills. Expertise can be defined in many ways; exceptional practice can be defined either in terms of relative proficiency in skills beyond minimal expectations, which is a norm-referenced testing task, or as mastery of specialized skills, which is a criterion-referenced or mastery-testing task.

Summary

In summary, licensure examinations differ significantly in six major areas, namely, (1) type of test; (2) test content; (3) administration; (4) scope of authority; (5) purpose; and (6) point of effectiveness. Differences in the type of test relate to testing strategy; ie., whether it is norm-referenced, criterion-referenced, or performance-based. Usually staff and time limitations force the test into an objective test mode which is machine scorable. The content of the licensing examination may be taken from the training program, from the observation of practitioners at work or from mixed content selected by professional test makers. Usually

examinations built on training content are made up by members of the profession who have gone to some form of professional institution. Test content coming from the field of practice usually contains items submitted by selected practitioners. Modes of test administration differ in that some tests are given by professional organizations at the end of a training program; come are commercially subscribed either at the point of application for a job or nationally through government agencies. The range of authority for a licensing examination may extend over state, nation, professional association, including either license to practice, accreditation, or special recognition. The purpose of the test may be to test people out for lack of competence or to test in to note special qualifications. Differences in the point of effectiveness range over the entry period or extended period of practice and the ability to be updated.

All of these six areas of difference must be considered in the design of a licensing examination for a given purpose and in the gathering of evidences of the quality of a given licensing exam. An analysis of today's practices in medicine, law, and education suggests that these professions have failed to adequately consider all of these six differences in the design of their current licensing examinations.