University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Licensure Testing: Purposes, Procedures, and Practices

Buros-Nebraska Series on Measurement and Testing

1995

Section Two: Overview of the Procedures for Developing a Licensure Examination

James C. Impara
University of Nebraska-Lincoln, jimpara@unl.edu

Section Two

Overview of the Procedures for Developing a Licensure Examination

James C. Impara

Buros Institute of Mental Measurements

There are a variety of strategies that may be employed in the development of a licensure examination. The following list of activities illustrates typical procedures. Depending on the needs and conditions of the particular occupation, certain variations in specific activities may take place or changes in the sequence may be appropriate. In addition to the procedures listed, many decisions will be made that may add activities. For example, the decision to use a computerized item bank or to enter into a computerized adaptive testing format will require procedures in addition to those described briefly below.

1. Conducting a job (or practice) analysis. Often practitioners in the occupation are surveyed to assess the nature of the job; the essential knowledge, skills, and abilities (KSAs) associated with the job; the extent that the KSAs are critical to performance in the profession for the purpose of protecting the public; and the extent that these critical KSAs are at the entry level of practice. An initial list of KSAs is often developed by a committee (perhaps supplemented by "shadowing" some practitioners and seeking additional insights from interviews with others) and a survey questionnaire is drafted and piloted. The pilot testing leads to expansion and development of the questionnaire that is then sent to a sample of practitioners. Responses may be analyzed by staff or a consultant.

2.   Developing test specifications based on the job analysis. The licensure board, or a test committee, determines the specific content dimensions and nature of the test by examining the job analysis (often assisted by quantitative analyses of the survey done by staff or a consultant). It may be possible to obtain a copy of the test specifications (or even the job analysis) of the tests currently being used in the same or related fields in other states or nationally and compare them with each other to assess the essential differences among the occupations of interest.

3. Making a decision about test development.   Strategies at this stage include: (a) develop an original test from scratch, (b) use a test already developed by another state or a national organization, or (c) attempt to collaborate with other states to develop a new test. Test development is an arduous and long-term project that has far-reaching implications and costs. If the occupation to be licensed is unique this may be the only available option. If original test development is the decision, the development process could be expected to take up to 2 years and it should be done with the advice of a test specialist/consultant.

Ideally, test development includes such activities as drafting original test items, reviewing the items by a testing committee, revising the items, pilot testing the items, assessing the psychometric properties of the items (DIF analysis) and the total scores (reliability and validity studies), refining the items based on the pilot test results, assembling the items into a final form (and having one or more alternate forms is desirable).

4. Arranging for test administration. The examination must be administered and scored. Various decisions need to be made (e.g., distributing the tests, employing test proctors, insuring test security, compliance with the ADA legislation) before the testing program can become operational.

5. Arranging for test scoring and data analysis. Reports need to be developed for the Board as well as for candidates who were tested. Reports to the Board often include various statistics about the test (e.g., item analyses, DIF analyses, reliability estimates, overall and subgroup score distributions). Examinees will need to know their test results, licensed or not, and the Board may decide to provide some diagnostic feedback to candidates who failed the test.

6. Setting a passing score. This is one of the most important aspects of the process and one of the most frustrating. There are several ways to set such scores, but only a few of these ways are considered by the testing profession to be defensible in the context of licensure. The nature of licensure precludes certain mathematical procedures—regression analyses that relate test score to future competency is virtually impossible; and arbitrary criterion-referenced ("70%" of the total score) and norm-referenced (scores above the "national average") methods may not hold up in court. Among the most often used methods is one reported by Angoff (1971). His method involves a panel of judges who independently examine each item and estimate the proportion of minimally competent examinees who will answer correctly. These proportions are averaged across panelists and the sum of the average proportions represents the cut score (often some adjustments are made to the derived cut score based on sampling error or measurement error or both). There are other similar methods that rely on

expert judgment and an examination of the test items, and there are some variations on the Angoff method.

7. Equating the test across test forms. The test form should be reconstituted by replacing all or most items for each administration of the test (for reasons of security—there will be candidates who are repeating the test who may be unfairly advantaged by having seen the items previously). This will require equating each different form of the test to some base form. There are several ways to accomplish this, but the decision about how it will be done is needed before the first test form is developed (the method used has implications for how the test is constructed).

The above model for organizing a licensure testing program is fairly standard for any testing program. Many important steps have been mentioned only in passing (e.g., conducting reliability and validity studies) yet these are critical steps and involve much time and energy if they are done properly. Section Two of this book describes the means for accomplishing these steps.

REFERENCES

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.; pp. 508-600). Washington, DC: American Council on Education.