The Ins and Outs of Creating High-Stakes Examinations

Introduction
As the name suggests, high-stakes examinations are vitally important to the careers of the people who take them. Certification and licensure are essential qualifications for some individuals to work in their chosen professions, and they provide the means for these individuals to market their expertise to consumers – the public or the employers seeking skilled workers.

A well-designed, high-stakes examination provides a solid foundation for a certification or licensure program, and a well-maintained program creates numerous benefits for the companies, organizations, or governments that sponsor examinations including:

- A brand identity that promotes the idea of protecting the public from harm and helps in attracting new professionals interested in taking advantage of the brand.
- An added revenue stream as candidates pay fees to take the examination or maintain their certified status.
- A cadre of outside people with a vested interest in distributing and maintaining examination-related products, freeing the sponsor of the burden and expense of providing such services itself.

This guide outlines many of the steps necessary to develop a credible examination. It is intended to help companies and organizations that are contemplating initiating a certification or licensure program to determine if they have the necessary support and resources for such an undertaking.

Key Terms

**Content validity** is the extent to which the content of the examination contains a balanced and adequate sample of questions representative of the knowledge and skills an individual needs for successful and competent job performance. It is essential that questions adhere to standards for clarity, accuracy, and fairness, with references and statistics providing essential documentation.

**Reliability** means that the scores individuals earn by taking the examination are dependable measures of candidate proficiency. Scores must be consistent regardless of when an individual was tested or which version of an examination was used.

**Defensibility** means that the examination is likely to withstand legal challenges from failing candidates and others who may claim the examination is flawed. The sponsoring company or organization can help to fend off such challenges by demonstrating that the examination was developed using a rational, thorough, and fair process, is content valid, and has reliable results, with a passing standard that is determined through the use of appropriate methods and seen by the profession as reasonable.
Psychometrician refers to someone with training in the field of measurement and evaluation, as well as extensive experience in assessing people’s abilities. The psychometrician is key to a quality high-stakes examination. He or she has the expertise to manage the development process so that each examination possesses the three hallmarks of a credible assessment instrument: validity, reliability, and legal defensibility.

Job/Task Analysis
The initial step in creating a high-stakes examination is a job/task analysis (JTA), also called a role delineation study. A JTA helps ensure that an examination is valid by identifying the abilities needed for successful job performance. JTAs generally are conducted in two phases. First, subject-matter experts (SMEs) define the primary knowledge and skills required of the certified professional. Second, a representative sample of practitioners reviews and validates the information to verify the accuracy of the SMEs’ view of the profession.

SMEs should be thoroughly knowledgeable of the profession and provide a cross-section of practice settings, geographical regions, ethnic backgrounds, and genders. The panel outlines the major performance domains – the principal areas of responsibility – involved in the profession, identifies the specific responsibilities or tasks associated with each domain, and lists the knowledge and skills associated with the successful performance of each task.

A survey of practicing professionals is used to validate the competencies outlined by the SMEs. Developing an appropriate sample of survey participants is critical to the JTA. The selected sample should represent both experienced and entry-level practitioners, as well as a variety of organizational settings and geographical regions. The sample should be large enough to maximize the ability to interpret the survey results definitively and make reasonable generalizations, especially regarding the differences in practice that may exist among respondents who represent various geographical regions, industry classifications, levels of education, and experience.

Surveys are conducted through the mail or via email, and the results are analyzed to determine how professionals in the field rate the competencies needed for successful job performance. Examination specifications, or blueprints, are created from these ratings, with more questions devoted to more important areas. Some of the skills that must be tested may not be appropriate for a traditional multiple-choice examination, so a simulation or practical examination may be developed to assess these decision-making skills.
Examination Development

Once the JTA is completed, the high-stakes examination is built to assess a candidate’s abilities in the essential job-related skills and knowledge. Through multiple-choice questions, candidates recall facts, apply specific knowledge to a given problem or situation, or reach an appropriate conclusion by analyzing or evaluating information. Other item types may be useful if it is necessary to assess certain skills or abilities.

A development committee workshop usually is held to write test questions, or items. Items should be written by people who have a significant level of experience in the profession and who represent different areas of expertise, geographical regions, and organizational settings. During the workshop, participants practice writing and reviewing items. Each item is reviewed and classified by at least three experts in a specialty area, in conjunction with the entire development committee. Items should reflect the knowledge level of minimally competent professionals. Multiple-choice items should be written so that the correct answer is defensible, and incorrect answers are plausible, but clearly not the best choice.

After editing for grammar and consistency and to eliminate any bias, the items are sent back to the committee for technical review. Those approved by the committee are normally placed into a field test that is administered to a sample of professionals in the field (or field tested as experimental items on the examination for existing tests). Following the field test, some items may be rewritten or discarded if they are found to be confusing or not useful in determining competence.

The development process should generate enough items to fill not only a single multiple-choice examination, but also multiple versions of the test, which prevent people from seeing the same questions if they have to take the test more than once. Item development and review can be accomplished quite well using online systems that enforce essential quality standards and maximize convenience.

Once items have been generated, they are stored in an item bank, a secure database that also contains background information on how each item was developed and its record of validation and use. This information provides the foundation for the quality and defensibility of the examination, while item statistics help identify poorly performing items in the future.

Follow-up workshops are often held to assemble the examination using the questions in the item bank. The examination is assembled according to the blueprint established by the JTA, and the passing point, or cut-score, may then be determined. A valid credentialing examination must have a defensible passing score that is based on the minimum competence required to perform proficiently in the profession or job. A criterion-referenced approach – considered by the testing profession to be the most defensible approach for setting passing points – relies on the pooled judgments of content experts who review each item on the examination to estimate the probability that a “minimally acceptable” candidate will answer it correctly.

Data Analysis

After each administration of the examination, candidate scores and statistics must be analyzed to determine if the test is functioning properly, as well as to improve its overall quality. An item analysis shows how well each question performs and indicates if a question is too easy or too difficult, if a question’s difficulty changes through time, and if a question’s right or wrong answer indicates how well
a candidate scores on the examination. Summary statistics of a test administration show the range and
distribution of scores.

As each new examination version is assembled, the content and difficulty level of the examination may
change, so each new version must be equated to maintain a constant standard of difficulty. Various
statistical methods can be used to accomplish this. Equating procedures work to maintain the passing
standard across multiple forms.

Combining examination score statistics with candidate demographic information also can yield useful
information, such as whether issues such as bias could affect scores or how individuals who take an
examination multiple times fare versus first-time candidates.

**Castle Worldwide, Inc.**
As one of the nation’s leading certification and licensure testing companies, Castle Worldwide, Inc., is
prepared to assist certification boards, government agencies, corporations, educational institutions, and
trade and professional associations with their high-stakes examination development and administration
needs. We have more than 25 years of experience in creating and delivering examinations and in helping
organizations build successful certification programs. Our psychometric staff includes nationally
recognized leaders in the field.

Please contact us with questions about the information contained in this guide or for more information
about the examination process and Castle’s capabilities.

**Castle Worldwide, Inc.**
900 Perimeter Park Drive, Suite G, Morrisville, NC 27560
Telephone: (919) 572-6880  Fax: (919) 361-2426
Email: info@castleworldwide.com
Website: www.castleworldwide.com