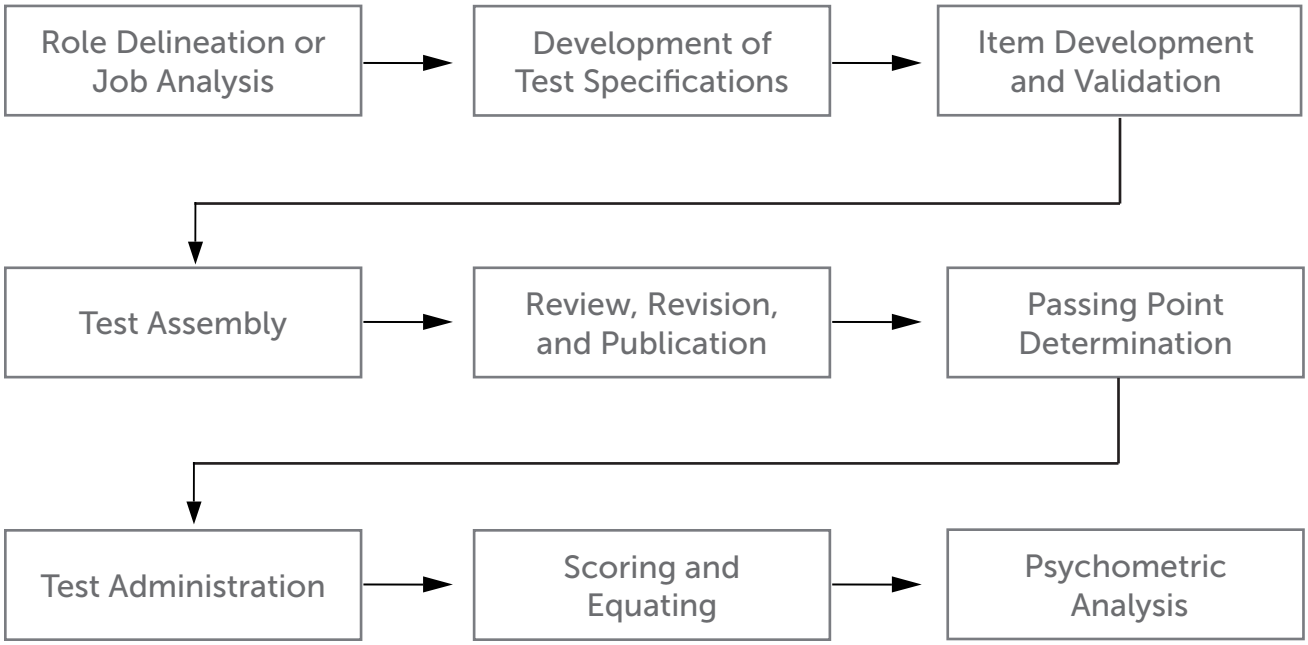
Castle
*Intelligent solutions for quality testing*

# Steps in the Construction of a Content-Valid Examination




## Role Delineation or Job Analysis

Before examination development begins, a role delineation or job analysis study is conducted. It determines the knowledge and skill benchmarks that define minimum proficiency for newly certified professionals in a field.

During the role delineation process, a committee of subject matter experts (SMEs) defines the domains of responsibility for newly certified professionals in the discipline. Typically, these domains are further broken down into tasks, and then into the knowledge and skills required to perform the tasks competently. The responsibilities identified by the SMEs are then validated through a survey of practitioners in which they review and rate the domains and tasks according to their importance, criticality, and frequency of performance.

Linking the validated knowledge and skill benchmarks to examination content ensures that it is valid. Content validation is how a test developer documents the competence to be inferred from scores.

## Development of Test Specifications

In the next step, the validation survey results are used to develop a weighted blueprint or plan for the examination. The specifications guide the item development and examination assembly processes and ensures that the examination represents the relative emphasis that is appropriate for the newly certified professional. Each domain and task's importance, criticality, and frequency is translated directly into the percentage of items that should be included in the examination for each content area.

## Item Development and Validation

All examination items are written by SMEs in the field. Each item writer is trained in writing, reviewing, editing, and validating questions. Each question is reviewed and validated by other SMEs and must have at least one verifiable reference. Each item is classified by content category, validated according to its appropriateness to the newly certified professional, and checked to ensure it is psychometrically sound and grammatically correct.

## Test Assembly

Each examination is created by selecting the appropriate number of items from each content area, as specified in the test blueprint. Statistics that describe item performance help in the selection of items.

## Examination Review, Revision, and Publication

The draft examination is reviewed by SMEs for technical accuracy and by Castle to ensure its psychometric integrity. Item performance data are available for items that have been used on previous examination versions. As examinations are administered, reviewed, and revised, item performance data and related statistics are used to remove or revise poorly performing items. With implementation of this and the preceding steps, it may be said there is evidence of the examination's content validity.

## Passing Point Determination

A valid credentialing examination must have a defensible passing score. The score that separates examinees who pass from those who fail must be based on the minimum competence required of the newly certified professional in order to protect the public. A criterion-referenced approach called the Modified Angoff Technique is often used to determine the cut score or passing point of an examination. This technique is currently considered one of the most defensible criterion-referenced methods available.

## Test Administration

Test administration procedures must ensure consistent, comfortable testing conditions for all examinees. For secure examinations, test administration procedures must address examinee admission into the room, seating, display of informational signs, security, and time allocation. Testing facilities must meet guidelines that ensure security, proper room size, ventilation, restroom facilities, handicap accessibility, and noise control.

## Scoring and Equating

The first step in scoring is key validation using item analysis statistics, followed by the application of the correct scoring algorithm, consistent with the psychometric model being employed and the design of the examination. Equating is a statistical process that ensures that different forms of an examination achieve comparability in difficulty; that is, examinees are not penalized if a form is harder than others or rewarded if a form is easier. Scaling scores as part of equating helps to ensure that scores on different forms maintain the same meaning. Accuracy and fairness are essential in this very important activity.

## Psychometric Analysis

The item statistics are reviewed to ensure quality and validity. Items with poor performance statistics are evaluated by SMEs prior to scoring. These items then are tagged for review at the next meeting. Scoring procedures include the computation of scores, following the established system of equating to ensure compatibility in scores over time and across different versions of the examination. Scoring procedures also include scaling to assist in the appropriate interpretation of the scores.