

Essential Elements of Certification

Purpose

This document provides an introduction to the testing process, especially as it applies to written and practical examinations for Crane Operators. This document defines terms such as licensure and certification, as well as reviews the concepts of validity and reliability. The document describes the process followed in the development of written examinations and some of the unique challenges faced in developing practical examinations. Hopefully it will assist in better understanding the underlying testing concepts which form the foundation of the certification examinations.

Credentialing

Credentialing is the process of granting a credential, that is, a designation which indicates competence in a subject or area. There are three forms of individual credentialing: licensure, certification, and registration. The usage of the terms is frequently interchanged.

Licensure

Licensure is the most restrictive of the three forms of credentialing and generally refers to a mandatory governmental requirement necessary to practice in a particular profession or occupation. The ultimate intent of licensure is to protect the public from incompetent practitioners. Licensure almost always requires assessment, which usually means an examination. Successful completion of an educational program and/or relevant experience is sometimes required. Licensure implies both practice protection and title protection, in that only individuals who hold a license are permitted to practice and/or to use a particular title. A few states and local governmental jurisdictions require crane operators to be licensed.

Certification

Certification is usually a voluntary process instituted by a non-governmental agency, such as the National Commission for the Certification of Crane Operators, in which individuals are recognized for their knowledge, skill and competency. The intent of certification normally is to inform the public that individuals who have achieved certification have demonstrated a particular degree of competency. As with licensure, certification normally requires testing and an evaluation of education and/or experience. Certification implies title protection as only those who are certified may use a particular title.

Registration

Registration is also used for title protection and normally requires individuals to apply for the title through the appropriate governmental or private agency. Registration is generally the least restrictive form of credentialing, most often used when public protection is less critical.

Essential Elements of Testing: Validity and Reliability

The psychometric principles underlying the testing process for the three forms of credentialing are the same. The authoritative source of professional technical guidance for the various participants in the testing process (i.e., test developers, users, takers, sponsors, administrators, and reviewers) is *Standards for Educational and Psychological Testing*. The purpose of the *Standards* is “to provide criteria for the evaluation of tests, testing practices and the effects of test use.”

The concepts of validity and reliability provide the foundation upon which measurement is based. While there is more to measurement than validity and reliability, many measurement concepts can be linked to an underlying issue of validity and reliability. The concepts of validity and reliability relate directly to the issues of fairness and accuracy that are critical to the candidates who seek a credential. In constructing certification examinations, every effort must be made to assure that they are valid, reliable and legally defensible.

Validity

As traditionally defined, validity indicates the extent to which an examination measures what it purports to measure. The common question “Is this a valid examination?” really cannot be answered with a simple yes or no response. It would be inappropriate to flatly state that an examination is valid. Rather, evidence is collected to provide support for the validity of the examination results. Validity is a matter of degree, which relates, in part, to the strength of the evidence provided. The most important principle of validity is that an examination is valid only for its intended purpose. It is the inferences regarding the use of the test that can be validated not the test itself.

Content validity is based on a concept of sampling; that is, taking a representative sample of the job itself. You try to determine whether performance on the test matches performance on the job, or on some part of the job. Content validity is established if the content of the test closely approximates the actual tasks or duties to be performed on the job, or the knowledge, skills, abilities and other characteristics required for successful job performance.

Reliability

Reliability refers to the degree to which the results of testing are free from errors of measurement. A test score for a particular person is an estimate of the person’s knowledge level in a certain subject domain. The estimate may or may not be an accurate indication of an individual’s true level of knowledge. Stated another way, reliability provides an indication of the degree to which results will be consistent over different forms of the same test on different testing occasions.

Legal Defensibility

Legal defensibility is the extent to which the process used to develop a certification program or examination, and the means by which it is administered and maintained, conforms to consensus standards and can therefore stand up to legal challenge.

Subject Matter Experts

Certification examinations are developed through the active participation of practitioners in the field, known as Subject Matter Experts (SME's), and the guidance and experience of a professional testing consultant or company.

The Subject Matter Experts define the context in which a certification program will need to operate and provide the content for the examination. The testing consultant provides the expertise in managing and facilitating the process.

Defining Purpose and Objectives

Once the decision to develop a certification program has been made, the certification body must follow a step-by-step process which begins by asking itself a number of questions, such as:

- Who will attempt to become certified and for what purpose?
- How should the public interpret the certification?
- How difficult should it be to become certified?
- What, if any, eligibility requirements should be established, i.e., education, experience, age.
- What knowledge and skill areas should be tested?
- How can candidates best demonstrate their competence?
- What kind of examination format would be most appropriate?
- Once certified, does one remain recertified indefinitely or is there a process for recertification?

Job Analysis

A Job Analysis builds the foundation for a content-valid examination. It identifies the tasks the practitioner (operator) does on the job and links test content to job requirements. A Job Analysis is used to determine:

- what tasks a person does on the job
- the importance of each task
- the frequency of a task
- the knowledge, skill or ability (KSA) required to perform the task

Information for the Job Analysis can be obtained through:

- research of professional literature, training manuals, etc.
- observation of practitioners at work
- focus groups of subject matter experts
- surveys of practitioners

The Job Analysis is used to develop the test specifications and the content outline.

Test Specifications/Content Outline

The test specifications define important attributes of the test. Test specifications include:

- a clear statement of the purpose of the test;
- a definition of the target population of candidates;
- the number of items on the test;
- the length of time allowed for candidates to take the test;
- the item types;
- scoring procedures;
- the content outline.

The content outline is the **Test Blueprint**. It outlines areas of knowledge or domains that the examination will encompass. Each of the content areas on the outline is “weighted” to reflect its relative importance. The foundation for the content outline is derived from the job analysis.

Examination Types

Certification programs use criterion-referenced tests, in which an individual’s performance is evaluated against a pre-determined set of standards.

Multiple choice items are most commonly used in certification examinations. Advantages include:

- Standardized administration;
- Ability to test various cognitive levels;
- Quick and accurate scoring;
- Cost;
- Historical experience;
- Research findings.

Examinations may also include the following types:

- True/False
- Matching
- Open-Ended
- Essay
- Oral Examinations
- Practical Examinations
- Portfolio Examinations

Item Development

Subject Matter Experts (SME's) are trained in item writing and item review by the testing company. Test items (questions) are written by SME's and linked to the knowledge areas listed in the content outline. Items are edited by test editors and entered into a data bank, known as an item bank. SME's review all items and decide whether to accept, revise or reject an item. The questions SME's ask during this review include:

- What knowledge, skill or ability is this question testing?
- Where does it best fit in the content outline?
- Is the item referenced to an authoritative source?
- Is it an important enough concept to be included on the test? i.e., not trivial
- Does it focus on one issue or problem?
- Does it reflect current crane safety procedures?
- Is all necessary information given/is there any unnecessary information?
- Is it clearly written, unambiguous and straightforward? i.e., no tricks, no obscure terms or words
- Is there only one correct answer?
- Are the distractors plausible/believable?

Pilot Testing

Accepted items are field-tested prior to inclusion on the examination. The results of pretesting indicate the difficulty of each question and provide information that enables SME's and test developers to determine whether items should be accepted, modified or rejected. Items meeting appropriate performance levels are approved for use on examinations.

Establishing the Passing Score (Standard Setting)

A psychometrician trains SME's, designated as judges, in the method used to arrive at an appropriate passing score. The SME's determine the probability of qualified candidates answering each question correctly. These values are summarized to establish the passing score.

Test Administration

Written examinations must be administered in a facility that meets appropriate testing standards, especially separation of candidates. Trained Chief Examiners/proctors administer the test under rigid security. Tests are returned to the testing entity for scoring. Test performance is analyzed prior to releasing scores to candidates. Reports are generated relating to candidate performance and are sent to the candidate. Results are entered into the item bank and into the candidate's permanent record.

Summary Test Statistics and Item Analysis

Results are analyzed to see both how the overall test performed, as well as how each item performed. Key factors reviewed include:

- Mean Score - average score on the test.
- Standard Deviation - a measure of dispersion of a set of test scores.
- KR-20 - a measure of reliability based on the internal consistency of an examination.
- Standard Error of Measurement - an estimate of the standard deviation that would be found in the score for a person tested repeatedly on the same test.
- P-Value - the percentage of candidates who get the item right. The difficulty level of an item determined by the proportion of the correct responses.
- Discrimination Index (Bi-serial Correlation) - the relationship between candidate performance on an item and their performance on the examination. Used to determine the extent to which performance of an item is related to performance on the total test.
- Distractor Analysis - the distribution of answer choices. Ideally each distractor should be selected by some of the test takers. Wrong answers should be more persuasive to less qualified candidates.

Below is a “live” examination question and the actual item analysis statistics generated in 2003 based upon 3,378 candidates.

According to ASME/ANSI B30.5, a signal person is required to observe power line clearances when the crane is within

- (A) 15 feet of a power line.
- (B) 10 feet of a power line.
- (C) a crane’s length of the limit of approach.
- (D) a boom’s length of the limit of approach.

Question Number: 35		P-Value: .77		R-Bis: .62	Item Number:
Response Options	A	B	C	D*	OMIT
Frequency (#)	115	556	115	2592	0
Frequency (%)	3	16	3	77	
Mean	50.88	55.62	57.2	66.17	

Practical Examinations

There are a number of terms often used interchangeably to mean practical examinations; such as performance tests, clinical examinations, performance demonstrations, skill assessments, etc. Practical examinations have the distinct advantage of being able to replicate specific work related tasks and require the candidate to perform those work related tasks as they would at a job site. Measuring an individual's skills and abilities through the use of a practical examination is an ideal situation. However, for a number of reasons, licensing boards and credentialing organizations are reluctant to use practical examinations. Among the reasons are:

- difficulty in handling the logistics of a practical examination;
- concern about meeting required testing standards of objectivity and uniformity;
- costs of a practical examination.

Selection of Practical Examination Tasks

Practical examinations are often selected as part of the testing requirements for a credential because the job itself requires a complex coordination of mental and physical skills for safe practice on the job. The opportunity to perform in a real or near-real situation is therefore a necessary component of the test. The skills selected for a practical examination must be based upon a job analysis which defines the skills required at the job site. The skills selected must also be a representative sample of skills required on the job. There should be a range of difficulty in the skills so that they are neither too simple nor too difficult. It is critical that the requirements for each candidate be the same in order to achieve the objective of fairness.

Standardized Testing Conditions

Standardization involves creating conditions that assure uniformity of the test with regard to administration, difficulty, scoring and testing conditions. With the crane operator examination, two important elements to be standardized are the test site and the testing equipment (the cranes). The test site must be laid out according to specific instructions and standards so that each candidate, regardless of the test location, is taking the test on a test site similar to other test sites. The cranes are required to be configured in a similar manner at all test sites.

Candidate Information

It is important that candidates receive complete information regarding the nature of the practical examination. A candidate handbook has been designed to provide a full description of the tasks that each candidate will be required to undertake, what constitutes acceptable performance and the time limits within which the performance must be accomplished.

Reliability of Practical Examinations

Provided the practical examination consists of representative work samples, validity is usually not a problem. It is the reliability of the exam that must be demonstrated. The reliability of the practical is based on standardized testing conditions and the consistency of the rater's judgements or inter-rater reliability.

* Factors related to the reliability of practical examinations include:

- relevant scoring criteria;
- technical quality of the instruments;
- inter-rater reliability - consistency between raters;
- intra-rater reliability - consistency of a rater over a number of candidates.

Scoring Criteria

During test development, a scoring system is developed by a test committee of subject matter experts. The relative importance of each task and each section of the exam is identified. The points are distributed first by the importance or weight of each section and then by the importance or weight of each behavior making up the section. The passing score is a critical factor related to test validity. The passing score is determined by measuring the difficulty level of each observed behavior.

Technical Quality of Rating/Scoring Instruments

The technical quality of the score sheet, rating scales or checklist is one factor influencing the reliability of rater judgments. Exam tasks must have tightly written and recording procedures including any time limits made as exact as possible in order to minimize opportunity for subjectivity on the part of the Examiner. The Examiner records the candidate's performance only. Actual scoring (grading) is done by a third party testing company.

Establishing Rater Reliability

The rating instrument cannot be considered reliable by itself. The raters must be trained to achieve a high level of agreement when using the examination. This is accomplished by having a rater training sessions in which raters score samples of work performance. During these sessions, scores are recorded, inter-rater reliability is measured and feedback is provided until a high level of agreement between raters is achieved. To be effective, training sessions need to be repeated periodically.

Two concepts of agreement are important in the scoring system. One concept is agreement across raters for an examinee or set of examinees. This concept can be thought of as one examinee and multiple raters: stability between raters.

A second concept is that of internal reliability, or the agreement within a practical examination for one rater. This is often called intra-rater agreement and refers to stability of judgements of an individual examiners.

Calibration

To maximize the extent of inter- and intra-rater agreement, examiners receive training, sometimes referred to as calibration. Calibration represents the degree to which several examiners or judges identify the same level of correctness for a given performance or the degree to which a single judge identifies the same level of correctness for the same performance over several examinees.

Reducing Subjectivity and Bias

In the field of performance testing, some subjectivity is inevitable. Examiner performance evaluations serve as a basis for important decisions such as certification, regardless of their reliance on subjective judgement.

Rater performance is a significant factor in practical examinations. The rater must be qualified and well trained. Rating forms should be behaviorally anchored so as to restrict the rater's range of observations. At the same time, the observed behavior must be worded so that the critical performance is scored and nothing important to competent performance is left out.

Summary

The integrity of the certification process is dependent on a number of key factors including:

- the thoroughness of the test development process.
- the quality of the assessment instruments, i.e., the written and practical examinations.
- consistent ongoing monitoring of performance statistics.
- the third party independence of the testing organization.
- continuous effort to maintain and improve the certification product and process.

The "proof" of whether a certification program is consistently meeting these standards can best be found in the accreditation process. The foremost accreditation body in the US is the National Commission for Certifying Agencies (NCCA).

The NCCA accreditation standards are developed and administered by experts in the field of competency assurance. Achieving accreditation provides objective evidence that an organization meets the level of quality represented within the NCCA standards. Because the standards are based on current guidelines and practices, accreditation can assist in demonstrating the fairness and appropriateness of certification programs in legal challenges. NCCA has embraced the mission of promoting quality practices and principles of fairness in competency assurance.