

# The Use of Psychometrics in NDT Certification Programs

Marvin TRIMM, American Society for Nondestructive Testing, Savannah River National Laboratory, Aiken, United States of America

**Abstract.** The NDT certification examination process must provide evidence that the candidate has sufficient skills and knowledge to correctly evaluate critical components for initial/continued service. But we must also remember the process has an impact on the livelihood of a candidate seeking certification. So the importance of the accuracy and fairness of the examination process cannot be overstated. We must also have evidence that examinations are related to the work to be performed, and be “a reasonable measure of job performance.” There are prescribed steps in linking the relatedness of the examination to the knowledge and skills required for a job. These steps lead to an examination that has been “validated” in that its content accurately measures the necessary knowledge and skill required for the job. One of the steps in the validation process must include a means to monitor examination performance. We must remember, the question of whether or not a certification examination is valid cannot be answered with a simple “yes” or “no.” But if you do, someone will always say, “Prove it.” Now that you went on record, can you “Prove it?”

An answer that every NDT Certification Body should be able to truly state is: “Our certification examinations have been developed using the appropriate methods to ensure that the examinations contain content that fairly reflects the knowledge and critical abilities required to effectively perform the jobs necessary for a particular Method/Level of NDT Practitioner.” Stated more simply, the answer might merely be: “Our exams adequately cover the defined scope of the job and fairly discriminates between those that can perform and those that cannot.” In today’s world, the Certification Body must be able to support this statement. Certification Body’s must be prepared to provide evidence that the appropriate methods were followed for ensuring that the examinations are “valid” for the NDT Method and Level for which the examinations were developed. This paper will discuss how Psychometrics (the science and technology of mental measurement) principles can be used to develop a NDT certification program that provides a process and necessary evidence that its examinations are “valid”.

## Introduction

First let me define Psychometrics: **Psychometrics** is the field of study concerned with the theory and technique of educational and psychological measurement, which includes the measurement of knowledge, abilities, attitudes, and personality traits. The field is primarily concerned with the study of differences between individuals and between groups of individuals. It involves two major research tasks, namely: (i) the construction of instruments and procedures for measurement; and (ii) the development and refinement of theoretical approaches to measurement. The first psychometric instruments were designed to measure the concept of intelligence. The best known historical approach involves the Stanford-Binet IQ test, developed originally by the French Psychologist Alfred Binet.

This measurement focus is on the NDE Certification process and more specifically the American Society for Nondestructive Testing’s process. NDE Certification Bodies have two

very important items to consider. First, because Safety is a major element of why we perform NDE, we must ensure that certified personnel have the skills and knowledge necessary to select and perform the appropriate NDE methods/techniques, correctly interpret all indications, select/apply appropriate acceptance criteria and report results to maintain the safety aspects of products. Secondly, successful completion of the NDE Certification process has an impact on the livelihood of a candidate seeking certification, the importance of the fairness and accuracy of the Certification process cannot be overstated. So as you can see, we must ensure public safety while providing a process that is fair and accurate. The US Equal Employment Opportunity Commission's *Guidelines on Employee Selection Procedures*<sup>[1]</sup> require evidence that tests are related to the work to be performed, and the US Supreme Court has held that tests must be “a reasonable measure of job performance.”<sup>[2]</sup> This paper will discuss how Psychometric principles can be used to ensure personnel have the necessary skills and knowledge to correctly apply the various NDE Methods while developing/maintaining a certification process that is fair and accurate.

The discussion of each psychometric component in this presentation includes *guidance* and a brief discussion of the *implications* for validity and reliability. Fairness is considered to be the sum of validity and reliability; that is, if a psychometric procedure promotes validity and reliability, it intrinsically promotes fairness.

In the US, The American National Standards Institute's ANSI-PCAC-GI-502, *Guidance on Psychometric Requirements for ANSI Accreditation*, was developed to provide guidance about compliance with ISO/IEC 17024 Conformity Assessment — General Requirements for Bodies Operating Certification of Persons to certification bodies interested in ANSI accreditation. It does not prescribe specific statistics that should be computed and displayed. Rather, it emphasizes *methodologies, procedures, types of analyses, and how they are applied*, as a basis for the accreditation standards in ISO/IEC 17024<sup>[3]</sup>.

There are prescribed steps in linking the relatedness of the examination process to the knowledge and skill required for a job. These steps lead to an examination process that has been "validated" in that its content accurately measures the necessary knowledge and skill required for the job. The question of whether or not a certification exam is valid cannot be answered with a simple “yes” or “no.” However, an answer that could be made by a psychometrician might be: “Our exams adequately cover the defined scope of the job.” To support this statement, the psychometrician must be prepared to provide evidence that the appropriate methods were followed for ensuring that the exams are “valid” for the job for which the test has been developed. As Certifying Bodies we must assume the role of the psychometrician when we are asked about the validity of our examination process.

To document validity of a certification exam, three assumptions are important:<sup>[2]</sup>

1. That there are certain critical abilities necessary for effective performance and those individuals who lack these abilities will not be able to perform adequately in practice.
2. That individuals scoring low on the exam lack knowledge underlying these critical abilities and will not be able to practice in a safe and effective manner.
3. That the exam can be designed to accurately identify the point at which the knowledge, skills, and abilities demonstrated on the examination are most indicative of the candidate's ability to practice in a safe and effective manner.

There are several steps in the psychometric process that are used to “Validate” our process, they are:

1. Job Analysis
2. Develop and validate items
3. Develop Examination Specification
4. Develop an exam

5. Establish a cut score
6. Evaluate Examination Performance
7. Revise examinations/Certification Process as necessary

Now that I have identified the steps, I would like to discuss in some detail each of these steps.

## **Job Analysis**

Conducting a job analysis is an essential first step in establishing the content validity of a certification process. A job analysis identifies the critical components of a job that are necessary for successful performance. In addition, a job analysis often lists the capabilities (i.e., knowledge, skills, and abilities) required to perform work tasks.

The first step in conducting a job analysis is to develop a broad outline of job activities. Many techniques can be used to develop this outline, including observing work situations, interviewing operators, and analyzing job-related material. After an outline of the job activities is identified, specific job tasks may be determined. In the NDE Industry, specific job tasks may identify the conduct of the various NDE Methods (i.e., perform Liquid Penetrant Examination, or Ultrasonic Examinations, etc.).

Discussions by Subject Matter Experts (SME) are held to identify specific job tasks and capabilities required for successful job performance. During these discussions, the SME's verify that the task statements developed are technically correct, unambiguous, and accurately reflect the job. After consensus is reached among the SME's that the task statements are adequate, the SME's identify the capabilities required to perform the job tasks. Identification of capabilities must be done on a task-by-task basis, so that a link is established between each task statement and requisite capability.

Within American Society for Nondestructive Testing (ASNT), generic NDE Job Task Analysis are performed by ASNT's Technical and Education (T&E) Council. The Method Committees members within our T&E Council are subject matter experts within that method. These committees develop what we call "Bodies of Knowledge". These Bodies of Knowledge are in the form of an outline and they provide the skills and knowledge necessary for each NDE Method by level. The ASNT's Bodies of Knowledge were contained in our certification documents (SNT-TC-1A and CP-189), but based on an ASNT Board of Directors Charge to the T&E Council, they were extracted from these documents and assembled in one US National Standard known as ANSI/ASNT CP-105 (*American National Standard ASNT Standard Topical Outlines for Qualification of Nondestructive Testing Personnel*).

It must be understood that the analysis performed by the ASNT Committees are generic to the specific method and typically does not include other elements that may be more industry or employer specific. The other elements that may emerge from a company or industry specific NDE Job Task Analysis should be added to items identified by ASNT's T&E Method Committees when you are developing a specific certification program for a specific industry or company. How these items are addressed is one of the many responsibilities of the responsible NDE Level III. These other elements may include employer/industry specific requirements which could range from plant/industry specific safety requirements to specific national codes or standards mandated by local, state, national regulatory agencies.

## **Develop and Validate Items**

Examination items (individual questions) are developed from the results of the job analysis so that examinations are representative of job tasks. Once the new items are written, they must go through a validation process, which includes:

- A. Attributing new questions to specific items identified in the job analysis (NDE Body of Knowledge). The purpose of attribution is to ensure that all questions in the question bank is linked to at least one important aspect of an inspector's job. Once you establish an examination data bank of questions, the certification body can construct an examination that will provide a demonstration of the knowledge and skills as they are used in the actual work.
- B. Analyzing questions for technical accuracy, style, readability, and possible bias to subgroups. Determine that the correct answer is the *best* answer and the answer is traceable to technical reference (Example: NDE Handbook, ASME Code, etc.), that the distracters (incorrect answers) are wrong, and that the question is free from bias with respect to race, gender, and culture.
- C. Reviewing items for difficulty level. Difficulty ratings should reflect what percent of competent inspectors, at the level being tested, should get the answer correct. Difficulty ratings will be assigned for new questions via table top analysis. Once questions have enough statistical data points, the question difficulty will be a product of the routine statistical analysis.

The continued relevance of questions that have been validated must be ensured through periodic reviews of the items by subject matter experts. Evaluation of questions should also be conducted through statistical analysis. Of particular importance are the difficulty index (the ratio of examinees that answer each question correctly) and the discrimination index (how well the question distinguishes between the more knowledgeable and less knowledgeable examinees).

## **Develop Examination Specification**

Test specifications demonstrate and document the job-relatedness of the examination. Examination Specifications (i.e. the categories of required knowledge such as safety, materials, discontinuities, and physics associated with NDE Method, etc.) are based on the results of a job analysis and reflects how often a task, knowledge, skill, or ability is needed in practice and how much impact it has on effective job performance. Simply stated, the SME's review the body of knowledge and assign a percentage (or number of questions) to specific topics or skills. These values will dictate the balance of the examinations as it relates to the Body of Knowledge topics/skills (Example: 10% of the MT Level II General Examination questions will address discontinuities).

## **Develop an exam**

Once the Specification has been established, the SME's will select the predetermined mix of questions (or use a computer program to randomly select the questions while complying with the examination specification). The selection (manual or computerized) is based on two factors: 1) the examination specification and 2) question difficulty factor. The questions selected must reflect the appropriate exam specifications that were identified by the job analysis as well as by statistical performance of the questions. This process will allow the Certification Body to generate multiple examinations that have proper concentration of questions and a difficulty factor that is consistent between the examinations to be administered. These two items are at the core of a fair and accurate examination.

## Establish a cut score

The cut score is defined as the minimum score required to pass an exam. Defining the cut score required for certification is one of the most important but difficult aspects of the validation process. The validity depends on whether the cut score accurately distinguishes between adequate and inadequate performance. Cut scores must be high enough to protect the public but not so high as to unnecessarily screen out qualified examiners.

A cut score study is critical because it addresses the difficulty of individual questions on an exam.

For example, an exam with more difficult questions should have a lower cut score than an exam with easier questions. The cut score is determined by judgments about test questions. Subject matter experts are used to specify the level of performance that should be required on the exam. There are many methods of conducting a cut score study (e.g., Angoff, Ebel, and Nedelsky)<sup>[4]</sup>. In order to be valid, a cut score study must be clearly documented and follow a structured process.

The following table is an example of the Angoff Method for 5 raters (A through E) and 10 items.

Each rater (a subject matter expert) rates each of the 10 items. Their collective average score for the 10 question exam is averaged to obtain the single cut score for the exam. The cut score for the example below is 70.1%.

	<b>Rater</b>				
<b>Item</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
<b>1</b>	.90	.85	.80	.80	.85
<b>2</b>	.90	.90	.90	.90	.95
<b>3</b>	.80	.60	.60	.70	.70
<b>4</b>	.75	.70	.70	.80	.60
<b>5</b>	.80	.55	.60	.55	.60
<b>6</b>	.60	.70	.65	.65	.70
<b>7</b>	.55	.50	.55	.55	.50
<b>8</b>	.75	.70	.55	.75	.70
<b>9</b>	.80	.90	.80	.80	.75
<b>10</b>	.55	.65	.50	.45	.65
<b>Rater</b>					
<b>Average</b>	.740	.705	.665	.695	.700

**Average Cut Score 70.1%**

## Evaluate Examination Performance

After an examination has been given, in addition to scoring each individual paper, it is important to evaluate the overall results. Were the scores higher or lower than expected? What was the distribution of scores? Do the results indicate that the examination served its purpose? If a similar examination was given before, have the changes incorporated in the present exam been useful? Were the new items effective? How can the items, old or new, be improved? Should the examination be changed in any way? Evaluation of results of an examination can be performed in many ways. Two commonly used methods are consideration of *examinee comments* and analysis of *examination statistics*<sup>[5]</sup>.

Examinees should be permitted to comment about individual items or about the examination as a whole. Such comments provide a subjective evaluation which is useful primarily as an index of the examinees perceptions.

The reviewer can use comments to estimate the general attitude regarding factors such as fairness of the examination or item, clarity of the wording, whether the time allowed was sufficient and whether the references provided were clearly printed and well organized. In considering comments, it is important not to attach too much significance to comments by any one individual. A consensus of the comments of the entire group of examinees will usually be more useful in revealing meaningful attitudes, errors or ambiguities. This does not mean that a single individual's comment should be ignored, but that it is more likely that an isolated attitude may not be shared by others.

Two general types of statistics are usually studied when evaluating examination results: *item analysis statistics* and *overall statistics*. Item analysis statistics are used to examine various attributes of the individual items on the exam. The factors that are most commonly evaluated for each item are:

1. Difficulty – the percentage of candidates correctly answering the item,
2. Discrimination – the correlation of those answering the item correctly with their overall score on the exam, and
3. Effectiveness of each response – the distribution of answers across the options.

The examination attributes that are usually measured are:

1. Number of examinees attaining each score,
2. Distribution of the scores, and
3. Mean, median and mode of the scores.

In addition, it is important to determine the reliability of the examination

Examination Statistics show the distribution of scores. Reviewing the examination statistics for each administration of an exam indicates how the exam and the candidates perform. George Wheeler and Dr. Larry Early have authored an excellent book (*Guide for Developing NDT Certification Examinations*) if you are interested in statistical specifics of this process. But I do not have time to discuss the details of the various types of statistics that could be used to evaluate examination performance.

Results reporting could be based on raw scores, percentage correct, or some form of the standard scores (z-scores: Indicates, for each raw score, how many units of standard deviation that score is above or below the mean of the raw scores, t-scores: Based on a mathematical conversion of the z-score to eliminate negatives and decimals, and scaled scores: scores using a scale other than the raw score or percent correct.). Score reports are a matter of certification agency policy. While a report provides a rank ordering, the candidates should not see this information.

The purpose of the certification program is not to reward the high scores, but to recognize any score above the pass mark as meeting criteria for earning certification. Results can be reported to the individual as pass or fail.

When a candidate fails, additional information on strengths and weaknesses across the topics of the exam should be provided. This feedback will allow the candidate to focus future studies on areas of weakness. An explanation of how the exam is scored must be provided in examination handbooks and materials so that candidates know in advance what to expect and how to interpret the scores and results.

In addition to the candidate that fails, one could argue that a successful candidate should also be given areas of weakness. After all, the successful candidate will be performing the work and we all agree that they should be aware of any weakness they may have in their knowledge and skill to apply the specific NDE Method. But the Certification Body must determine to what extent feedback will be provided to examination candidates.

## Revising the Examinations and/or Certification Process

One of the most important elements of the certification process is maintenance. We all know that our industry is very dynamic and its technology is advancing on a daily basis. For a certification process to be fair and accurate as we have discussed, we must attempt to mirror the every changing technical elements of our industry.

As I have just identified, the candidate comments and statistical analysis will identify needed revision to examinations and our certification process. Many comments will only address the technical issue of the specific question, but some comments address outdated techniques and equipment. We must address both types of comments.

We must understand that our NDE Certification process will always be a dynamic. We typically react well to minor changes in the methods and techniques that are currently address in our certification process. But the inclusion of new product forms, materials, and/or NDE techniques/equipment tend to have a slower reaction time. Some lag in the reaction time must be expected because of the validation process we follow. But we must attempt to keep close to the leading edge of technology and our applications of that technology.

Our certification systems, must reevaluate the state of the art on a real frequent cycle. One way this does occur is by the current evolution of our Job Analysis (Bodies of Knowledge). As you know they are revised periodically (4-5 years) to maintain pace with our evolving industry. The certification bodies must react to these changes and continue to develop and validate test items to reflect changes in our Job Analysis. But it is the responsibility of the Certification Body to push their program to ensure it reflects and evaluates the knowledge and skill required to perform current NDE applications.

## References

- [1] US Equal Employment Opportunity Commission, 1978. Part 1607–Uniform Guidelines on Employee Selection Procedures. *Federal Register*, Vol. 43, No. 166–Friday, August 25, 1978.
- [2] National Organization for Competency Assurance, 1996. Certification, A NOCA Handbook, Anne Browning, Alan Bugbee, Jr., and Meredith Mullins (eds), Washington, DC.
- [3] American National Standards Institute, 2004, ANSI-PCAC-GI-502, Guidance on Psychometric Requirements for ANSI Accreditation, Washington, DC 20036
- [4] Impara, James. 1995. Licensure Testing: Purposes, Procedures, and Practices. Lincoln, NE: Buros Institute of Mental Measurements.
- [5] Early, Larry Allen, Wheeler, George C. 2005, Guide for Developing NDT Certification Examinations, Columbus Ohio, The American Society for Nondestructive Testing, Inc.