Introduction

The process of developing and administering professional certification exams comprises several discrete components, each requiring high levels of rigor and expertise, facilitated by robust information technology systems.

Each of these components—job analysis and standard setting, item development, item analysis, test design and exam construction, and test publishing—pose distinct challenges to organizations involved in professional credentialing.

Despite the critical role played by professional credentialing in modern society, professional organizations and others involved in the credentialing process often struggle to establish and maintain rigorous processes in the creation and publishing of certification exams.

This is the third in a three-part series of reports on the credentialing exam development process, based on data from a survey sponsored by Prometric.

The first report provides an overview of current practices in the exam development process. That report shows that many organizations rely on informal practices and improvised solutions in order to meet the challenges of exam development.

The second report focuses on question (item) development and exam design. That report focuses on common challenges and limitations identified by survey respondents, including budget and time constraints, difficulty recruiting and motivating subject matter experts, and inadequate IT systems, among others.

This third and final report takes a closer look at key factors of the exam development process: job analysis, standard setting for pass-fail decisions, and exam publishing. These components are critical...
to ensuring the construct validity and overall integrity of credentialing exams. Yet, the survey data show that these components are often not executed completely or efficiently.

**Job Analysis and Standard Setting: Current Practices and Challenges**

Professional certification and licensure is intended to ensure that everyone allowed to practice a profession meet or exceed minimum standards of knowledge and skill. In order to fulfill this goal credentialing exams must be based on some sort of job or practice analysis. Ideally, job analysis is conducted by skilled researchers with considerable experience assessing the knowledge, skills and abilities (KSAs) required to successfully carry out the duties and activities of a particular job or profession.

However, finding the personnel and other resources to conduct a proper job analysis can be a challenge, one that increases with the professional level, complexity, and level of independence and creativity of the job under analysis. It is easier to identify and catalogue the KSAs required of an assembly line worker than a neurosurgeon. It takes more effort and skill, and perhaps more data sources (interviews, focus groups, documentary evidence, etc.), to properly analyze these more complex professions.

Furthermore, the data generated through the job analysis must be translated into meaningful and appropriate exams. Test content should have a clear and explicit link to the tasks identified in the job analysis as being critical. In other words, the credentialing exam must have construct validity relative to the job or profession in question. This can be challenging in an environment where resources are limited. It takes time and money to ensure that the content of the exam matches the tasks and duties of the profession.

Construct validity is also a concern when setting performance standards to determine passing scores on exams. The number of correct answers required to pass the exam and receive professional credentialing has to be determined by analyzing the content relative to the profession and difficulty of the exams.

Job analysis and standard setting become even more challenging and critical when the profession in question is evolving at a rapid pace. It can be difficult to keep up with the state of the art in fast moving professions. Yet, failure to do so, means that the construct validity of exams can erode more quickly over time.

**Job/Practice Analysis – Uses and Techniques**

Job or practice analysis is the systematic and thorough collection and analysis of job-related information (duties, tasks, activities, etc.) for a particular purpose, such as establishing job hierarchies and career ladders, determining compensation, improving organizational efficiency and service quality, ensuring compliance with equal employment opportunity and non-discrimination statutes, and determining the necessary and sufficient knowledge, skills and abilities (KSAs) for professional credentialing.

Practice analysis requires information from individuals who are knowledgeable about the profession, such as practitioners, supervisors, managers, and educators. Additionally, non-human sources of information can be critical. These include job descriptions, quality control analyses, and textbooks, among others. Common techniques employed to conduct practice analysis include the task inventory, the professional practice model (PPM), the critical incident technique (CIT), functional job analysis, and comprehensive practice analysis, to name a few.

Underlying these established approaches are a number of data collection techniques, including observation of practitioners, interviewing knowledgeable individuals, administering work diaries to small samples and questionnaires to large samples, auditing records, and conducting focus groups with subject matter experts or practitioners.
Job Analysis and Standard Setting

Despite the critical need for job analysis in the development of credentialing exams, a high percentage of organizations do not carry out this step of the test development process. When asked simply if their organizations have developed such analysis, 30% said no.

Respondents were subsequently asked what types of research studies their organization has conducted to demonstrate the construct validity of scores utilized in pass-fail decision processes. Responses to this question suggest that job analysis as a core component for ensuring that credentialing exams reflect the jobs and professions they are supposed to be testing is even less widely used.

Two-fifths of respondents (40 percent) did not specify any type of research that could demonstrate the construct validity of their test scores. Another 6 percent did not know, while 13 percent referred generally to conducting job or practice analysis, but did not specify any techniques, tools or approaches.

These data indicate that likely more than half of respondent organizations do not conduct any type of job analysis or any other research or analysis that would assure the construct validity of their tests.

<table>
<thead>
<tr>
<th>Information Sources</th>
<th>Evaluation Criteria for Test Results</th>
<th>Limited Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject matter experts</td>
<td>Cut score</td>
<td>General claim of job analysis 13%</td>
</tr>
<tr>
<td>Surveys</td>
<td>Modified Angoff</td>
<td>Don't know 6%</td>
</tr>
<tr>
<td>Focus groups</td>
<td>Standards</td>
<td>None specified 40%</td>
</tr>
<tr>
<td>Documentation</td>
<td>Bookmarks</td>
<td>None specified 40%</td>
</tr>
</tbody>
</table>

The relatively high percentage of organizations that do not use any practice analysis or other research analysis reflects the considerable constraints on time, money and talent (see the second report in this series for more detail). When asked to identify the biggest challenges their organizations faced in regard to their test development and content management process, many pointed to the difficulty of keeping their tests current with changes in the profession.

One respondent noted that the biggest challenge was: “…staying current with the practice. Just ensuring that the content is matching what is current in practice.”

Another respondent associated the challenge of staying current with the considerable amount of time it takes to create and validate test items. His biggest challenge was: “Time to write and test items. The body of knowledge in finance changes rapidly. We did not have enough time to write, in particular because of the changing body of knowledge.”
The high percentage of survey respondents whose organizations do not perform any practice analysis or evaluation for identifying qualified and unqualified practitioners is cause for concern. The pressures in terms of limited time, money and expertise are understandable. But, the very integrity of professional certification is potentially at risk without proper procedures for generating and validating test items and results.

**Test Publishing**

Test publishing comprises both the process by which the structure of an exam is developed and the process by which it is delivered. It does not refer to just the final delivery or administration of the exam. However, confusion and uncertainty exist among respondents as to what precisely the term refers.

Almost half of respondents said that the word “publishing” is adequate or it is simply the term the industry uses. Most of the rest said the term was either not appropriate, or understood it to refer narrowly to printing, distributing or administering exams.

Some of the respondents who understood the term narrowly said that the advent of electronic delivery of the exams made the term “publishing” imprecise, while others said that it is still appropriate.

**Exam Publishing Challenges**

Approximately three of every four respondents (74 percent) identified at least one challenge in the test publishing process. As a group these respondents identified an average of 2.1 challenges they faced in this critical component of the test development process.

By far, the most common challenge was the long lead time required from completion of exam construction to publication. Almost a third (32 percent) said this was a major challenge, mainly because it required longer planning horizons, limited the opportunities to make late changes, and created uncertainty and risk.

***Standard Setting Techniques***

Even with job or practice analysis, one has to take the next step to translate that analysis into a test and then establish criteria for evaluating test results to determine whether someone has passed or failed. In order to preserve the integrity of the profession, and legal defensibility of the credentialing process, the exam items and the interpretation of results must have high levels of construct validity. A few common techniques for establishing pass-fail criteria are briefly described below:

- **Cut score (contrasting groups method)** – A cut score is a point on a scale in which scores at or above the point are in a different category or classification than scores below the point (e.g. pass/fail). One technique for establishing the cut score involves first determining what constitutes “qualified” for qualified practitioners in a given field. Then the items on the test are administered and scored. Two groups of candidates with scores that represent “qualified” or “unqualified” are rescored in two separate distributions. The point at which the two distributions intersect is the location of the passing or cut score.

- **Angoff’s method** – The passing score is calculated from an estimate of the probability of a borderline candidate answering each item correctly. After a discussion and consensus of the nature of a borderline candidate, each judge on a panel independently estimates the probability that a borderline candidate will answer each item correctly. Judges’ assessments for each item are averaged to determine the probability of a correct response for that item. Then, each probability assigned to an item on the exam is averaged to obtain the pass point. The Angoff does not use exam performance data.

- **Bookmark method** – Item difficulties are calculated from exam performance data. Items are arranged from least to most difficult. Each judge on a panel selects the most difficult item a borderline candidate would be likely to answer correctly and a “bookmark” is placed at that location. Performance data (item p-values and candidate performance distribution) is then given to the judges. Judges then discuss the impact of the collective average of “bookmarks” or pass points, and are subsequently given a chance to change their bookmark. The number of items up to the final bookmark for all judges are averaged to determine the pass-fail point.
Typical of respondent comments was this one: “The timing of it, our vendor takes about 6-8 weeks to develop some exams, if we want something to hit the test centers, we have to plan. If something happens the time line is too long.”

Another respondent noted that publishing lead times require a tradeoff in other areas of the test development process. “The only challenge is the time it takes, about 3 months to go through the publishing process. If it were decreased we could spend more time on item development.”

The second most commonly cited challenge of the test publishing process is security, mentioned by 14 percent of the respondents. For most of these respondents no other challenges were mentioned, suggesting that for them security, was the overriding concern.

Two of these respondents associated the security concern with their organization’s use of printed exams. As one respondent said, their challenge was, “maintaining the security of it [the exam] during the printing process.”

While printing of exams may add to security and confidentiality concerns, they are not the only source of risk. The following section describes the numerous sources of security threats, and the measures that respondent organizations have taken to address them.

Exam Security
Security is a critical component and central concern throughout the test publishing process. Security concerns begin in the earliest stages of test development when subject matter experts start generating items for inclusion in exams.

Even for small and tight-knit professional communities in which senior practitioners are respected and trusted, potential leakage of proposed test items is still a concern. On the other end of the test publishing process, there are numerous potential security threats related to unauthorized use of reference materials during the examination, and tampering with exam results.

Throughout the entire test development and publishing process there are considerable security risks related to the reproduction or acquisition of exam material outside of the authorized chain of custody. Security breaches fundamentally threaten the integrity of exams and the credentialing process as a whole.

Furthermore, because of the considerable investment of time, money and expertise to develop test items and construct exams, security breaches can be very costly. One testing company in Canada estimates that it takes...
approximately 18 months to complete the exam development process. If a security breach were to occur in the latter stages of development it could cost millions of dollars, and delay the credentialing process.

As a group, survey respondents indicated awareness of security concerns throughout the test publishing and delivery process. Some respondents said that it was only the their vendor’s responsibility (16 percent) and some said the information was confidential (4 percent). The remaining respondents identified on average 2.5 security measures related to published exams, and a total of more than 16 different security measures.

Among the measures identified, using exam proctors was the most common (identified by 40% of respondents), followed by the required presentation of photo identification (often two photo IDs) by the test taker (30%). The third most frequently mentioned security measure is physical security (locks, metal cabinets for paper exams), which was mentioned by 21 percent of respondents.

Other security measures included encryption of electronic exams (19%), secure shipping of paper exams (16%), video monitoring and recording of test takers (16%), secure servers and networks (15%), and prohibition of personal items such as phones, jackets, etc. in the exam room (13%), and password access (12%).

This last measure – password access – relates to the exam development and construction process, as well as to the administration of exams. The same is true for non-disclosure agreements, which could apply to item developers and other subject matter experts involved in the creation of the exam, and to testing center personnel. Indeed, one respondent explicit referred to the requirement that item developers sign non-disclosure agreements and that they be sworn to delete test material from their personal files.

Some respondents also mentioned the prohibition against printing hard copies of exams. One respondent mentioned that his organization was about to migrate to using only electronic exams as a security measure. Others also mentioned the prohibition against use of email to transmit exams, favoring instead password protected file transfer protocol (FTP) for sending exam files.

Interestingly, among the other security measures identified were the analysis of exam results to detect suspicious response patterns, and the use of “secret shoppers” to ensure that testing centers fulfill their security commitments. Additionally, two respondents mentioned restricting the frequency with which an individual could take a particular exam, and two others mentioned the use of electronic whiteboards for performing calculations rather than writing on paper which might be carried out of the exam room.

In all, these data suggest that survey respondents are cognizant of the importance of exam security. Still, the most frequently identified measures (use of proctors and photo identification of test taker) are among the easiest to implement. The stakes are extremely high for credentialing organizations. Security threats can erase the investment of large sums of money and the product of months of labor invested in test development efforts. Even more concerning, security breaches that remain undetected can undermine the integrity of the credentialing process itself.
Conclusion

The integrity of professional credentialing exams relies on thorough practice analysis, careful standard setting, and efficient and secure test publishing procedures. The survey results reported here indicate that credentialing organizations as a group confront distinct challenges to the execution of these critical components of the test development process.

Unsurprisingly, time and money are the main constraints. On the front end, many credentialing organizations do not conduct thorough practice analysis or have specific standard setting procedures. On the back end, many organizations struggle with long lag times between compiling exams and publishing them, making ongoing changes and updates difficult.

Some of these challenges are driven by limited budgets and the reliance on traditional, informal practices in the test development process. Others, particularly in test publishing, are the result of cumbersome and inefficient processes. These constraints result in exams that reflect only imperfectly the requirements for professional certification. They threaten to undermine the construct validity of the exam.

This is the third in a three-part series of reports on current practices in the development of professional certification exams, based on survey research sponsored by Prometric. The first report provides an overview of current practices in the five phases of test development: item development, item analysis, test design, job analysis, and test publishing. The second report focuses on item development/analysis and test design. This report focuses on job analysis, standard setting and test publishing.

Find Out More

Learn more about our Test Development and Delivery Solutions by visiting www.prometric.com or by calling toll-free 1-855-855-2241.