Introduction
The proliferation of professional testing, licensure and certification programs in recent decades reflects the increasing reliance of modern society on expertise in critical fields from teaching to engineering, medicine, law and more.

For employers, licensure and certification of professional staff provides a way to communicate in a crowded marketplace their commitment to excellence. For employees, individual licensure and certification provides portable credentials that support employment mobility in an increasingly fluid job market. For customers, whether businesses or consumers, certifications provide assurance that the plumbers, electricians, doctors, lawyers, engineers, software developers, real estate agents, pilots, and others have been approved to apply their knowledge and skills by an organization intimately involved in building and maintaining the repositories of information and techniques that lie at the core of each profession.

The importance of professional licensure and certifications, and the exams that support them, should mean that the methods and processes by which tests are created are highly refined, uniform and rigorous. However, the first report in this series shows that informal, improvised mechanisms and processes are commonly used throughout the exam development process by a substantial proportion of organizations involved in professional licensure and certification.

This report takes a closer look at: (1) two critical components of the exam development process, including item (or question) development and test design; and (2) the needs and challenges of certifying organizations to successfully complete them.

Item Development: Current Practice
Test items are the building blocks of all exams. Items comprise more than just a question. They include additional instructions on presentation, response processing and feedback to the test taker, according to the IMS Global Learning Consortium, the non-profit consortium dedicated to developing interoperability.

Prometric Survey on Test Development Practices
To understand how professional testing licensure and certification organizations meet their needs, and the areas where they experience both success and difficulty, Prometric commissioned a survey of professionals responsible for managing their organization's certification process.

Telephone interviews were conducted with 82 individuals between February 25 and March 18, 2011. The survey instrument included a mix of constrained response and open-ended questions.

Respondents represented organizations of varying size, and from a broad range of industries and professions, including accounting, construction, education, healthcare, information technology, law and telecommunications.

What is an item?
"The smallest exchangeable assessment object... An item is more than a 'Question' in that it contains the question and instructions to be presented, the response processing to be applied to the candidates, response(s) and the feedback that may be presented (including hints and solutions)."

Source: IMS Global Question and Test Interoperability Assessment Test, Section and Item Information Model.
standards and adoption of practice standards for distributed learning. In this report, “item” and “question” are used interchangeably.

Constrained-response types of questions (those that offer the test taker a menu of options from which to choose the correct answer), are by far the most common type of questions used by respondent organizations. In contrast, so-called constructed-response types of questions, which typically take the form of short answers or essays, are much less commonly used, mainly because of the challenges of establishing and implementing consistent scoring criteria, and the time involved to score.

The relative popularity of constrained response questions is that they are quicker and easier to score consistently across questions and across test takers. Nonetheless, constrained response questions pose other challenges. The very ease of processing test results for constrained response questions – especially the archetypal multiple choice question – makes them challenging to develop. There must be unambiguously only one correct answer, which consequently requires the utmost conceptual and grammatical precision in writing such questions.

Furthermore, because the question presents the test taker with the correct answer among a set of incorrect answers (“distractors”), additional development effort is required to minimize the ability to “work” the question by eliminating incorrect choices in the hopes of increasing the odds of guessing correctly. Consequently, as the one respondent organization points out, this item type “requires a great deal of time to construct effective questions, especially ones that test higher levels of learning.”

What types of items do your exams utilize?

- Multiple choice 88%
- Multiple choice (exclusively) 39%
- Short answer/essay 22%
- Hotspot 11%
- Drag and drop 10%
- Simulation 10%
- Match 9%
- Fill in the blank 7%
- True/False 5%
- Scenario 4%
- Calculation 1%
- Other 4%
- Don’t know 4%

Glossary - Question types

There are numerous types of test questions, which have even more variations. At the most general level, there are two types of test questions: constructed response (or, “subjective”) questions and constrained response (“objective”) questions. Examples with advantages and limitations are listed below.

**Constructed response questions** usually refer to short answer and essay questions, but also include scenarios, simulations, calculations and practice.

<table>
<thead>
<tr>
<th>Type</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fill in the blank – Shortest form of constructed response (if possible answers are not presented to test taker)</td>
<td>+ Especially useful for language vocabulary and concept definitions</td>
<td>+ Less suited for assessing comprehension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Possibility of unanticipated, but correct answers</td>
</tr>
<tr>
<td>Short answer – Usually requires an original sentence or paragraph in response to question</td>
<td>+ Easy to construct</td>
<td>+ Need to ensure that questions do not have more than one correct answer</td>
</tr>
<tr>
<td></td>
<td>+ Minimizes guessing</td>
<td>+ Scoring is laborious</td>
</tr>
<tr>
<td></td>
<td>+ Student must know answer vs. recognizing answer</td>
<td></td>
</tr>
<tr>
<td>Essay – One of the longest forms of constructed response (and requires at least multiple paragraphs of original text)</td>
<td>+ Easy to construct</td>
<td>+ Limits amount of material tested, decreasing validity</td>
</tr>
<tr>
<td></td>
<td>+ Minimizes guessing</td>
<td>+ Expensive to score</td>
</tr>
<tr>
<td></td>
<td>+ Test taker can demonstrate ability to organize knowledge, be original</td>
<td>+ Scoring is laborious</td>
</tr>
</tbody>
</table>
Despite these challenges, constrained response questions are by far the most commonly used by survey respondents, and the multiple choice question is the most commonly used sub-type, with a substantial majority of respondents (88%) reporting its use. Almost 4 in 10 (39%) said that they used multiple choice exclusively. Adding those who said they use true/false questions (a form of multiple choice, with only two answer choices), 44% use multiple choice question types exclusively.

In contrast, less than a quarter (22%) used the short answer and essay types of question. Another 10% of respondents used simulations, and 4% used scenarios. Notably, some of these respondents suggested that the multiple choice question was the mechanism used to test understanding of their simulations and scenarios.

Satisfaction with Item Types Currently Used
A substantial majority of respondents (59%) indicated that they would like to use different or additional question types from the ones their organizations currently use.

Notably, respondents whose organizations use multiple choice questions exclusively were less likely to want to change than those using multiple choice as one of a set of question types (53% and 63%, respectively). This suggests that respondents with experience using various item types may be more comfortable experimenting with additional question types, or that they believe that using multiple choice questions solely is inadequate for testing in the fields they represent.
Many of those who use only multiple choice questions seemed confident in their current approach, saying they will “stay with what we have”, or “multiple choice works for us.” One of these respondents pointed to the challenges of including new items by noting: “We’ve been looking at short essay questions and matching. It just depends on price to get those implemented and the reliability of doing them.”

Still, more than a majority of both groups expressed interest in using other types of questions, suggesting that barriers exist that are preventing them from adopting the question types they prefer.

**Limitations Preventing Adoption of New Item Types**

In order to identify possible barriers to adopting an expanded set of question types, respondents were asked, without specific prompting, to identify the limitations that prevent them from adopting new item types.

More than three-quarters of the respondents (77%) identified at least one specific limitation to adopting new question types. The remaining 23% said they faced no limitations, that they were either satisfied with the question types they were using, or that they could adopt new question types whenever they might choose.

By a wide margin, the most commonly identified limiting factor was cost, reported by 41% of survey respondents. Technology constraints (22% of respondents) and time (15%) rounded out the top three limiting factors for adopting new question types.

Several respondents identified the need to find a vendor or the lack of vendor capability (9%), and the lack of in-house expertise (6%).

One respondent whose organization uses only multiple choice questions observed: “If you go to either free response or performance, you need to develop scoring guidelines. They [the responses to these types of questions] have to be scored by judges. It introduces a lot more cost into the process, which is okay, if justifiable. I see it as a fairly big undertaking. Then you have to revalidate the whole thing.”

Another respondent tied cost and time together, noting: “The biggest [limitation] is time to development. It seems like when we look at new items it takes 3 or 4 years to create and pretest before they go live. It’s a cost issue. Is it worth the money?”

For the 22% of respondents who mentioned technology as a limitation, most associated this limitation with cost, while others cited challenges integrating testing technology needed for new item types with existing systems.

Several mentioned the limitations of either in-house programming staff or their vendors. According to one respondent: “Our third-party vendor can’t handle delivery of new types of questions or it becomes too costly.” Another asked: “Is the item supportable by the vendors? We might have a great idea, but our third-party vendor might not be able to support [it].”

---

<table>
<thead>
<tr>
<th>What types of items would you like to include in the future?</th>
<th>Use multiple choice and other types</th>
<th>Use multiple choice exclusively</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different/additional types</td>
<td>63%</td>
<td>53%</td>
</tr>
<tr>
<td>Same as current</td>
<td>37%</td>
<td>47%</td>
</tr>
</tbody>
</table>

---

What current limitations prevent you from adopting new item types?

- Cost: 41%
- Technology: 22%
- Time: 15%
- Vendor: 9%
- Expertise: 6%
- None (satisfied): 23%

“The biggest [limitation] is time to develop. It seems like when we look at new items it takes 3 or 4 years to create and pretest before they go live. It’s a cost issue. Is it worth the money?”
A few respondents focused on technology for item banking, and for one of these the size of the item bank did not justify the investment in technology. According to one, “We’d like to use computer testing, but you need a bank large enough to do it.”

Among the 23% of respondents who said there were no limitations to adding questions, their responses typically focused on the fact that they already offer a wide array of item types, that they lacked interest, or that other question types offered limited utility. One respondent whose organization offers “over 40 question formats” noted: “We simply respond to our existing stakeholders, whatever their demands [may be].” Another respondent whose organization uses only multiple choice and essay questions said: “I don’t think there are any limits. I don’t believe other item types add significant value.”

**Limitations to the Item Development Process**

The most common limitations to the respondent’s item development process relate to writers: recruiting them, coordinating them (especially when they are geographically dispersed), and training them. Technology and item authoring tools are also a common limitation.

**Recruiting writers**

Recruiting writers was the most common limitation identified by respondents, with almost a third (30%) saying this made the item development process more difficult.

Many of these (12% of all respondents) also said that training writers to produce good questions was a challenge, particularly because of the rigors of the item writing and review process. One respondent summarized the challenge as: “making sure we have good item writers, so that in turn they’re producing high-quality items which can make their way through the process. Then at the end it doesn’t fail.”

Other respondents noted that subject matter experts (SMEs) often have difficulty dedicating the time and energy to item writing, especially if they are being asked to volunteer their time. As one respondent noted, the challenge is “…getting our SMEs to follow the instructions and make timely submissions. ‘Getting them to show up. ‘Getting people to agree without being paid."

A few respondents noted that their particular fields pose unique challenges, either because the field is advancing rapidly, or because the SMEs do not have the language skills and precision required for writing robust test items. As one person in the technology industry asked rhetorically: “How many programs can a person work on at one time? Technology may be moving faster than the ability to keep up and revise the test.” This sentiment was echoed by another respondent who said his organization’s main limitation was “item development for emerging practice regimens; the literature is just emerging.”

The building trades face a different problem: “being able to find SMEs that can write good questions. They’re construction guys.”
Item Authoring and Banking Systems, Geographically Dispersed

Writers

Inadequate technology to support the item writing process is a major limitation of respondent organizations. The lack of adequate item authoring tools was the second most frequently identified limitation of respondents’ item writing process, mentioned by 22% of respondents. Several of these respondents also noted the need for item banking tools (8%). These issues were particularly challenging for respondents who used geographically dispersed item writers (12%).

“In our current system [we] don’t have the ability to do online reviews”, observed one respondent. “If someone authors it, it’s in the item bank. They can’t log in and review the item. They need a PDF [Adobe Acrobat file] or physical piece of paper for review.”

Other respondents explicitly related the need for electronic item authoring tools to the geographically dispersed pool of item writers. “Right now”, noted one, “[the limitation is] with software. It’s just a little difficult to bring in people from remote locations. ...there are software limitations, and we’re working on ways to solve those issues.”

Similarly, another noted that the main limitation is “use of technology for remote submission. Multiple platforms cause incompatibility.”

Face-to-face Writing Sessions

Notably, some respondents emphasized the challenges of having geographically dispersed item writers, but did not regard technology as a necessary tool to overcome this challenge. These respondents seem to be committed to bringing writers together for face-to-face item writing sessions, either because that’s how they’ve always produced items, or because they firmly believe that productivity is enhanced this way. Such sessions, however, impose two constraints: place and time.

One respondent said the main limitation is: “getting the groups together to do the writing of questions."

Others said: “getting content experts together", and “the travel people incur and location". Another noted: “bringing everyone together and coordinating people", adding, “I held workshop teleconferences in Europe, Australia and Singapore — scheduling [is a limitation] due to time complexity.”

Challenges Finding Item Writers

As noted above, recruiting writers is the most common limitation of the item writing processes of respondent organizations. But, what specifically about recruiting writers poses the greatest challenges?

Authoring Systems and Item Banking

The terms “authoring system” and “item bank” are sometimes used interchangeably. However, IMS Global, the non-profit test interoperability consortium, defines them as two distinct concepts.

An authoring system, according to IMS Global, “allows item authors to create new items, to edit existing items imported from conforming item packages and to export items into new or updated item packages.”

An item bank, on the other hand, is “a tool for managing collections of items, their meta-data, and any associated usage data.”

IMS Global’s specifications state that: “A conformant item bank system allows item bank managers to import and export collections of items from item packages. Item bank systems must not alter the items” [emphasis added].

“Though a given tool may combine the features of an item bank system with an authoring system, to be a conformant item bank system it must still be capable of importing, managing and exporting collections of items without modification of the associated data.”
Finding people with appropriate expertise is the most frequently experienced limitation, as reported by 28% of respondents. The next most frequently mentioned limitation is finding experienced writers who also have the availability to dedicate to the item writing effort.

Several respondents emphasized the fact that writing test questions represents an additional responsibility that is secondary to SMEs’ main occupation. This was succinctly summarized by one respondent who observed: “You are asking of their time and expertise on a product that is not their prime job.”

This issue is compounded for organizations that rely on low compensation or volunteer labor, as this respondent noted: “We don't pay them. At least we pay them lodging and travel, and a very small honorarium of $50 a day. Well, just getting them to volunteer period, they’re busy, they’re professionals.” This was echoed by another respondent: “The payment for item writers is not sufficiently motivating to get the best talent.”

Commitment to Producing Quality Items

Beyond the challenges of recruiting writers with appropriate expertise, who have limited time for work that is often insufficiently compensated, numerous respondents emphasized more intangible qualities of commitment, natural ability, and even imagination that make good item writers – and that are in short supply.

One respondent noted the difficulty of finding people “with time and interest and the stick-to-it-ness to do [the job].”

Another elaborated: “Finding item writers that have a knack for it. Volunteers are easy to get, but what you want is good, quality items. Some individuals have it, some don't. Some have a good feel for writing multiple choice questions; their number of rejects is low. Others don't have it, so they create a lot of items, but they don't survive review.”

This view was shared with another respondent who said the challenge is “to continually find new people, find someone you like who sticks around for a long time. Its hard to find a good item writer; you can never train people to be a good item writer.”

Perhaps these last two perspectives overstate the issue, treating commitment and attention to quality as lack of innate ability. Still, they underline the substantial challenge of finding the complete item writer, one who has subject matter expertise, strong command of language, experience and understanding of what makes a good question, time to dedicate to the item writing endeavor, and commitment to quality.

Conclusion

Individual test items are the building blocks of any exam. The types of items used, the processes and systems by which they are produced and evaluated, and the qualities of the people writing those items
determine the robustness of the final product, the exam. The survey results in this report, show that the main challenges confronting professional certification organizations mainly center on cost and funding, which often cascades into other issues related to technology and personnel.

However, cost is not the root cause of all the other challenges identified by survey respondents. Some respondent organizations could realize efficiencies by breaking out of traditional, informal practices that entail reliance on volunteer labor and face-to-face meetings for group item-writing sessions.

Multiple and diverse pressures for currency in often rapidly developing fields, continuous need to replenish the ranks of item writers, reduction in the number of rejected test items, easier and more rapid item editing and feedback, and maintenance of large stores of items in test banks – even global diversity, make necessary ongoing assessment of licensure and certifying organizations’ overall approach to their item development and test design processes.

This is the second in a three-part series of reports on current practices in the development of professional licensure and 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. This report focuses on item development/analysis and test design. The third report focuses on job analysis and test publishing.