# Table of Contents

- Overview
- Flowchart
- Job/Content Analysis
- Test Specifications
- Item Writing
- Item Review
- Exam Review
- Cut Score
- Scaled Score
- Equating
- Statistical Analysis of Questions
Overview

The ASQ Certification Board has oversight responsibility for the exams, and its top priority is to ensure the integrity of the exam development process. To ensure that ASQ certification retains its status as a hallmark of excellence, the Certification Board uses continuous improvement processes and formally reexamines each certification program on a five to seven-year cycle. This ensures the continued relevance of the exams in the global marketplace.

ASQ certification exam development relies heavily on the efforts of ASQ certified members. During the course of a year, approximately 300 subject-matter experts (SMEs) are involved in some phase of exam development. Volunteers travel to ASQ headquarters in Milwaukee, WI, for two-day workshops to engage in exam-development related activities.

Workshop activities include development of a Body of Knowledge (BoK) or writing exam questions, and ensuring the validity of an exam through a variety of review processes. In return, the volunteers broaden their knowledge and skills through networking with peers from a variety of services and industries. They also come away with the knowledge that they have contributed to the development of many areas of the quality field.

The ASQ Certification Board has mandated that anyone involved in the development of an exam must refrain from helping certification candidates prepare for a test, either formally or informally, for a period of two years after their last involvement in the process.
ASQ EXAM DEVELOPMENT

5-7 Year Review of Existing Product

Job Analysis Workshop [12 SMEs] → Job Analysis Survey → Test Specification Workshop [12 SMEs]

New Certification Product

Camera Review Proofing of Exam [2-3 SMEs] → Exam Review Workshop [12 SMEs] → Item Review Workshop [12 SMEs] → Item Writing Workshop [16 SMEs]

Cut Score Workshop [12 SMEs] → Administer Exam

Item & Statistical Analysis → Score Exams and Notify Candidates

Overall Time Frame: 12-18 months  Workshop Length: 15-18 hours  SMEs Across Workshops: 78-79
Job/Content Analysis

The foundation of a valid, reliable, and legally defensible professional certification program is a properly constructed job/content analysis study. ASQ maintains a rigorous schedule for updating its Body of Knowledge (BoK) that is consistent with the Standards of Educational and Psychological Testing (AERA, APA, & NCME, 2014), and the ISO 17024 standard (sections 8.2, 8.4, 8.5) for credentialing agencies. ASQ utilizes two valid empirical methods for updating all its content:

**Job Analysis:** Through the job analysis process, the tasks performed on the job and the knowledge needed to perform them are identified and validated. The content outline of an examination is then linked to this empirical description of practice, creating a framework for an examination that is job related and content valid.

**Curriculum Study:** The curriculum validation study is used for certifications that do not have a formal functional job linked directly to them. This method is intended to provide candidates with the competencies to meet industry expectations as operationalized through the certification standards and industry groups as well as providing the knowledge to support technical skills. The curriculum study is a vital formative evaluation that will provide data to identify and inform any program content modifications.

Every new exam program must be approved by the ASQ Board of Directors as well as the Certification Board before it can be developed. Often, an ASQ Division supports the program and is the body that approaches the Board with marketing and research information, which can take up to 2 years to collect, in order to gain the approval. Once approval has been given, the job analysis process can be initiated.

For existing exam programs, the body of knowledge is updated approximately every 5-7 years. When this occurs, the process is completed again in order to determine what has changed in the industry for that exam program over those 5-7 years.

Whether a program is new or has been available for many years, the starting point for the exam development process begins with a job analysis. A survey is developed from the Job Analysis and is used to identify the skills and knowledge areas currently being used in the field to be tested.

**Job Analysis Committee (12 SMEs)**

To create an appropriate survey instrument, a job analysis committee made up of volunteers is appointed by the sponsoring ASQ division or the certification chair and committee of the exam. These volunteers are ASQ members, are already certified, and work in the area to be tested. ASQ volunteers who serve on the committee are known as subject-matter experts (SMEs).

The SMEs for the job analysis committee meet for two days, and their primary goal is to identify typical job responsibilities and the knowledge required to perform their job. The result of this meeting is a survey that asks respondents to rate each knowledge statement in terms of importance and frequency:

*How important is this task or knowledge?*

*How often is this task performed or knowledge used?*
Test Specifications

Survey
Once the survey has been approved by the job analysis committee, it is sent to a sample of certified or otherwise qualified ASQ members who either work in the area to be tested (as identified by job title) or supervise employees who perform the tasks identified in the survey. The survey is sent to a sample of individuals who actively hold the certification and all individuals have the same chance of being select in order to not skew the results of the survey.

Test Specification Committee (12 SMEs)
The committee includes some members of the job analysis committee and other qualified subject-matter experts who, together, represent a wide demographic of the industry and service spectrum.

The test specification committee meets for two days, and its primary task is to review the results of the task and knowledge statements in the job analysis survey to identify the content that should be included in the new Body of Knowledge (BoK).

These tasks and knowledge areas are then assembled into an outline format with topics and categories that logically group to form the BoK. Once the major topics are identified from the job analysis, and the BoK outline is created, the SMEs develop explanatory subtext that clearly describes the particular content in each subtopic. The number of questions and the time limit for taking the exam are also determined at this workshop using data collected from the survey results.
Item Writing

Item Writing Committee (12-16 SMEs)
The next step is the actual writing of the exam questions (items). This task is accomplished by a group of 12-16 certified SME volunteers during a two-day item writing workshop. The SME volunteers are given extensive training in a variety of exam development issues, including how to avoid writing trick questions and how to develop thought-provoking questions and answers.

The SMEs are then assigned to small groups where they write test questions on specific areas in the BoK. The processes used in this and other exam development workshops are driven by standards that are internationally recognized for the development of assessments. In keeping with these standards and as a means of ensuring the validity of the exam, each question must be linked to a book from a list of references for that certification. Each question developed at this workshop must have the following characteristics:

- Only one key (correct answer) that must have team agreement
- A reference to support the key
- A rationale for the item, which explains what the item is testing, why the key is correct, and why the other options are not correct
- A classification to the most detailed level of the BoK

Once a question writer has finished writing a complete test question, other members of the team review it for completeness, accuracy, and appropriateness. The item is then reviewed by ASQ Test Development staff. After approval by multiple reviewers at the item writing workshop, the exam item is entered into the ASQ exam bank as a “raw” test question.

The questions produced will be reviewed by another group of SMEs at an item review workshop before they can be used in an exam. The BoK is finalized after the item writing workshop to ensure questions can be written to each area of the BoK, and that there are adequate references to support the newly written questions.
Item Review

Item Review Committee (12 SMEs)
The next stage in the exam development process is to convene a panel of 12 subject-matter experts/volunteers who meet for two days to review the raw questions.

Prior to the workshop, the questions are sent to the committee members in an exam-like format. The committee members answer the questions as if they are candidates sitting for an exam. As they review the questions, they are invited to make comments about them and submit their comment sheet to the ASQ Certification Assessment Specialist who compiles everyone’s comments into one document. This document is used during the item review workshop as a guide for the group discussion.

The committee members verify the references and the BoK classifications, and most importantly, agree that there is only one correct answer to a question. The additional intensive processes of review, revision, and rework are designed to ensure that all language ambiguities have been eliminated and that all questions have been phrased and presented as clearly as possible. The questions are also reviewed to ensure that they are not biased in favor of any particular industry.

Last, the questions are reviewed again by the ASQ Test Development team.
Exam Review

Exam Review Committee (12 SMEs)
The next phase in the process is the exam review workshop in which subject-matter experts are asked to participate by reviewing the exam as a whole.

Prior to the workshop, the questions are sent to the committee members in an exam format. The committee members answer the questions as if they are candidates sitting for an exam. As they review the questions, they are invited to make comments about them and submit their comment sheet to the ASQ Certification Assessment Specialist who compiles everyone’s comments into one document. This document is used during the exam review workshop as a guide to the group discussion.

The committee members verify the references and the BoK classifications and again agree that there is only one correct answer to a question. They review the exam as a whole, determining if any questions overlap in content or key other questions, as well as, determine if any questions should be replaced with another question on the exam form. At the end of this exam review process, each question on the test has been reviewed and approved by dozens of qualified professionals. The result of the workshop is one complete exam that will be used as the pilot exam.

The exam is reviewed one last time by the certification exam committee virtually as a final check.
Cut Score Committee (12 SMEs)

The final phase in the process is the cut score workshop where a panel of 12 subject-matter experts are convened to determine the passing score for the exam. To be legally defensible in the U.S. and to meet the Standards for Educational and Psychological Testing (AERA, APA, NCME; 2014), a cut score cannot be determined arbitrarily; it must be empirically justified. That is, ASQ cannot merely decide that the cut score will be 70% correct. Instead, a study must be conducted to determine what score will best differentiate the classification of examinees in terms of competent vs. not competent.

The panel’s first task is to set the performance standard for the exam. To do this, they use consensus as they determine a set of characteristics that they expect a minimally qualified or “borderline” candidate may have in relation to the topics in the BoK. These characteristics can be found on asq.org as the “Minimum expectations” for all the existing certification exam programs. Once that list of characteristics is developed, the subject matter experts use it as a guide.

ASQ then uses the Modified Angoff Method, which involves the subject-matter experts estimating the proportion, or percentage, of candidates in the target group that they think would respond correctly to each item. This method takes two rounds of ratings with the goal being to reduce variability in participants’ judgments and converge on values within a target range of 25 points. Participants generate their first round of ratings individually, then discuss their ratings as a group question by question, and adjust their individual ratings to create their second round rating. The final cut score is obtained by calculating the grand average of participants’ ratings across all items in the test.

In addition to the Modified Angoff Method, ASQ also uses the Beuk Method for comparison. C.H. Beuk, a leading researcher and author in educational measurement, realized that standard setting “is only partly a psychometric problem.” That is, determining the accuracy and reliability of a testing instrument and its results (psychometrics) does not automatically establish the actual pass/fail point of a given exam. For the Beuk Method, the subject-matter experts are asked to provide their estimate of what passing rate should be expected for the exam after they have provided their estimates of the difficulty level for each test question for the Modified Angoff Method.

At the end of the workshop, the panel will have produced two recommended cuts based off of the Modified Angoff Method and the Beuk Method.

The exam pilots and the ASQ Test Development team collects and reviews statistics of each question verifying that the question performs well prior to grading the candidates.

The summary of the subject-matter experts’ combined estimate of the overall exam difficulty as well as the question statistics is presented to the Technical Advisory Committee (TAC) as the recommended cut point for the exam. Once that raw cut score point is established by TAC approval, it is converted to a scaled score (550), which becomes the minimum score necessary to earn certification in that BoK.
Scaled Scores

Why are scaled scores used instead of raw scores?
Multiple exam forms, based on the same BoK, are active in each administration. We recognize that each of these test versions can prove to be more difficult or easy than the test on which the cut score was set. To be fair to the candidates and consistent in maintaining the standard of performance from one test to another, we analyze the performance of the exam and adjust the cut-point on the raw score (the actual number of questions answered correctly) of the test.

If raw cut scores were reported, different passing scores could be shown for each version of the test, resulting in confusion. Instead, we use scaled scores to allow comparisons to be made between various versions of the test. A scaled score is a conversion of the number of questions answered correctly. The scaled scores range from 200 to 750 and reflect your performance level, while taking into account the difficulty level of the test.

The following is an example of a scaled score conversion for two versions of a test. This example is for a hypothetical 100-question test:
In the example, test A was slightly more difficult than test B, so the individuals who took test A were not required to answer as many questions correctly as the individuals who took test B. In the comparison table, the scores of 68 for test A and 70 for test B reflect the same minimum level of performance required to become certified. The scaled score of 550 is used to represent the minimum level as it is reported to the examinees.

The use of scaled scores for reporting purposes also helps you determine your score in relation to the minimum standard of performance for a test. Since the standard of performance is represented by a scaled score of 550, you can determine your performance compared to that standard. Likewise, if you took a test at two different administrations, your two scaled scores can be used to determine how much your performance has improved between the two administrations.

It is important to note that those who pass the test will not receive a scaled score because those candidates have met or exceeded the standard of performance.

<table>
<thead>
<tr>
<th>Test A</th>
<th>Test B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Raw Score</strong></td>
<td><strong>Scale Score</strong></td>
</tr>
<tr>
<td>65</td>
<td>530</td>
</tr>
<tr>
<td>66</td>
<td>540</td>
</tr>
<tr>
<td>67</td>
<td>540</td>
</tr>
<tr>
<td>68</td>
<td>550</td>
</tr>
<tr>
<td>69</td>
<td>550</td>
</tr>
<tr>
<td>70</td>
<td>560</td>
</tr>
<tr>
<td>71</td>
<td>560</td>
</tr>
<tr>
<td>72</td>
<td>560</td>
</tr>
</tbody>
</table>

Although the raw cut score is established for a specific number of questions answered correctly under a BoK, the scaled score is what is reported to the candidates. This scaled score allows adjustments to be made for exam difficulty on subsequent forms of the test, while maintaining a scaled score of 550. The scaled score of 550 is the minimum standard of performance for all ASQ certification exams.
Equating

How do we adjust for the varying difficulty level of exam forms?
The goal of ensuring that two versions of the same exam have the equivalent degree of difficulty is achieved through a process known as common item equating. ASQ selects a set of questions from the previous exam and embeds them in the next exam. This set of questions, called equaters, is a kind of mini-exam in that the questions are representative of the previous exam’s difficulty level (some easy questions, some hard, some in the middle) and cover areas of the BoK proportionately. ASQ then develops the rest of the test with different questions, some new and some previously used. This way, ASQ can administer new tests and still maintain the established standard of performance.

The common items (equaters) between the two tests are analyzed after the new test is administered to determine whether the new test is more or less difficult than the previous version. If the analysis of the equaters determines that the new test is more or less difficult than the previous test, the cut score is adjusted so that the standard of performance is maintained over time.

For example, if one version of a test is administered in March and the mean score of the candidates is 80, and another version of that test is administered in October and the mean score of the candidates is 75, does that mean that the test administered in March was easier than the October test or were the candidates who took the test in March better prepared than those who took the test in October? Before determining whether there was a difference in the tests and before making any adjustments to the cut point based on those differences, more information is needed about the two candidate groups. To gather that information, comparisons are made between the performances of the two groups on the common items (equaters) in the two tests.

The following is an illustration of how an analysis of the common items can determine whether the October test was more difficult.

<table>
<thead>
<tr>
<th>March Exam</th>
<th>October Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Exam Mean</td>
<td>Total Exam Mean</td>
</tr>
<tr>
<td>Score = 80</td>
<td>Score = 75</td>
</tr>
<tr>
<td><strong>Is there a difference in the tests or a difference in the groups?</strong></td>
<td></td>
</tr>
<tr>
<td>Common Items Mean</td>
<td>Common Items Mean</td>
</tr>
<tr>
<td>Score = 20</td>
<td>Score = 20</td>
</tr>
</tbody>
</table>

Since the two groups performed equally well on the common items (the mean score for both administrations on the common items is 20), it can be concluded that the two candidate groups were equally well prepared to take the test. Therefore, the expectation is that these two candidate groups would perform the same on the tests overall, i.e., display the same total test mean score if the two tests were equivalent in terms of difficulty. Since the overall mean score for October was lower than the mean score for March, it can be concluded that the test administered in October was, in fact, more difficult than the test administered in March.

As a result of this analysis, the cut point for the October test is adjusted to offset the effects of being a more difficult exam. Using this method ensures that both tests will fairly assess the candidates’ abilities while maintaining a consistent scaled score of 550 to pass.
Statistical Analysis of Questions (items)

Although ASQ goes through great lengths, with the assistance of SMEs, to ensure fair exam questions, questions may not perform as well as expected. Because of this, a statistical analysis is conducted on a regular basis to identify exam items that do not perform as well. Once identified, these items are reviewed by the appropriate members of the Certification Board as well as the ASQ Certification Test Development staff. If a question is determined to be unfair or inaccurate, ASQ marks the question a "score all". This means all examinees will be given credit for that question on the exam.