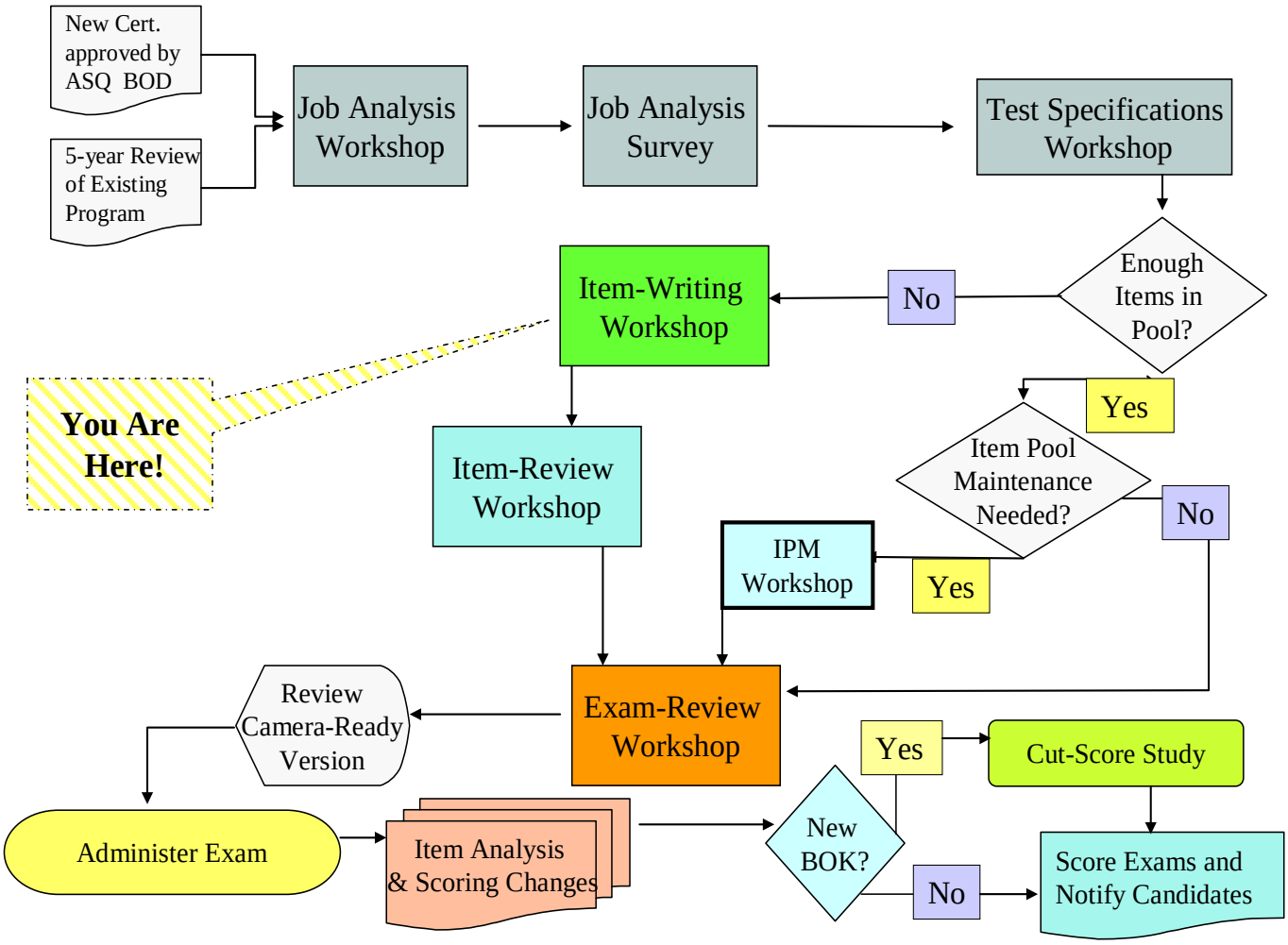
Welcome

to the

ASQ Item Writing Workshop

ASQ
The Global Voice of Quality™

STEPS IN TEST DEVELOPMENT



Why *Multiple-Choice* format Questions?

The advantages of multiple-choice questions can be clearly illustrated by looking at two item types that were used before multiple-choice questions.

- **Fill-in-the-blank or completion-type questions**

  Example: "A well-prepared audit checklist _____."

There are two reasons that this type of completion question doesn't work: 1) often insufficient direction or information is provided; 2) very different answers could be argued as acceptable.

- **True-false questions**

Example: "A product's reliability is directly linked to its purpose and use. T ___ F ___"

Although these types of questions *suggest* that there is no doubt about the correct response, they are, in fact, open to interpretation and the intent of the statement can be misread. In other words, they invite "absolutism" by the reader. True-false questions also give the candidate a 50% chance of being correct just by guessing.

Multiple-choice questions

A multiple-choice question with four stated options provides the candidate with: 1) greater direction, 2) only a 25% chance of guessing correctly, and 3) when devised with care, a more thought-provoking task.

Example:

Generally an auditor has the most planning and scheduling freedom for which of the following types of audit?

(A) Internal
(B) External
(C) Product
(D) Process

Why do we use "Which of the following"? Why not just ask "What"? Or why not ask "Which *one* of the following"?

"Which of the following" is a style often used in testing because it limits the allowable answers to just those listed in the four choices. Asking "What" can result in the candidate reading more into the question than was intended. At the same time, "Which *one* of…" is too restrictive and doesn't allow for plural answers when necessary. "One" is the "assumed" subject in most items.

Multiple-Choice Terminology

(or Why Do Test Developers Talk the Way They Do?)

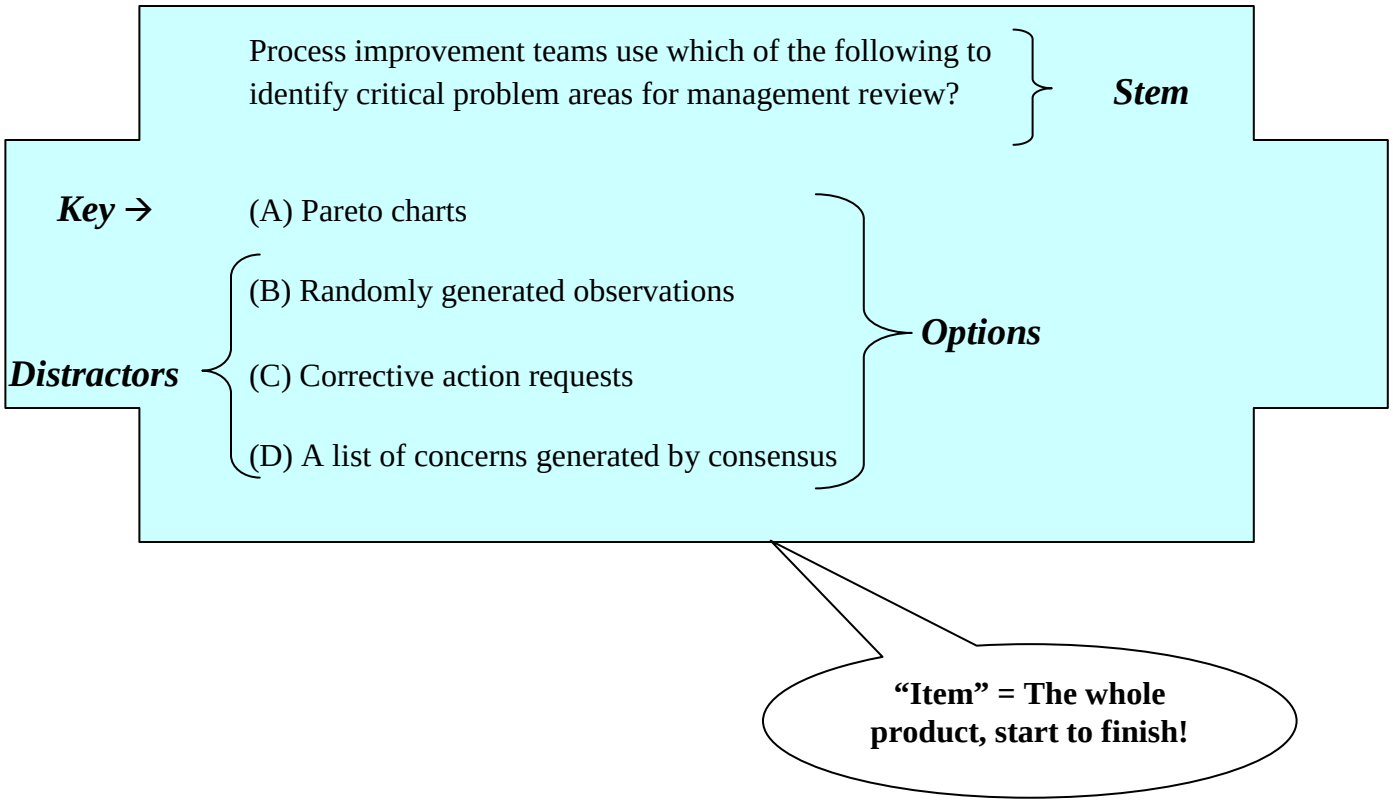
Stem: The initial part of the item in which the task is described in the form of a question, a direction, or an incomplete statement

Options: All of the choices in an item

Key: The correct answer – the answer that will get the candidate credit

Distractors: The *in*correct options (also known as "foils")

Item: The test question in its entirety, including the question and all of the answers, right or wrong



Types of Multiple-Choice Items

Various types of multiple-choice formats can be used, depending on the knowledge, skill, or ability to be measured. When constructing a test question, the item writers should use the format that provides the most direct, effective, and precise meaning.

A **CLOSED STEM** item has a question mark at the end of it.

> Which of the following is the most important factor when a continuous improvement team is selected?
>
> (A) Identifying staff members who are independent of the target process
> (B) Appointing only certified quality improvement associates
> (C) Ensuring that one of the team members is in a management position
> (D) Permitting the team leader to select compatible people

An **OPEN STEM** item is an incomplete sentence in which the options themselves complete the sentence. It is essentially a fill-in-the-blank question with the blank *always* at the end.

> Auditors must keep in mind that observations from any form of audit are the result of
>
> (A) the auditor's own opinions
> (B) sampling of a population
> (C) information from a hostile environment
> (D) careful judgment in balancing sensitive political issues

For this type of item to be effective, the information presented in the stem must be sufficient to enable a knowledgeable reader to know what he or she is looking for in the options.

"Auditor observations are" would not work as a good open stem.

Rank Order Format

This format is a good way to test the candidate's ability to put a process in an appropriate logical or sequential order in the most efficient or most economical manner possible.

Rank order, from first to last, the following time-management steps.

1. Make a to-do list
2. Assess personal time management skills
3. Create an action plan
4. Set goals

(A) 2, 1, 3, 4
(B) 2, 4, 1, 3
(C) 4, 2, 1, 3
(D) 4, 3, 1, 2

Caution:
If you present the options with a different first step for each option, you're really only testing the first step in the process.

The following information is contained in each body of knowledge (BOK).

Six Levels of Cognition
based on Bloom's Taxonomy (Revised)

In addition to ***content*** specifics, the subtext detail also indicates the intended ***complexity level*** of the test questions for that topic. These levels are based on the Revised "Levels of Cognition" (from Bloom's Taxonomy, 2001) and are presented below in rank order, from least complex to most complex.

Remember
(Also commonly referred to as recognition, recall, or rote knowledge.) Be able to remember or recognize terminology, definitions, facts, ideas, materials, patterns, sequences, methodologies, principles, etc.

Understand
Be able to read and understand descriptions, communications, reports, tables, diagrams, directions, regulations, etc.

Apply
Be able to apply ideas, procedures, methods, formulas, principles, theories, etc., in job-related situations.

Analyze
Be able to break down information into its constituent parts and recognize the parts' relationship to one another and how they are organized; identify sublevel factors or salient data from a complex scenario.

Evaluate
Be able to make judgments regarding the value of proposed ideas, solutions, methodologies, etc., by using appropriate criteria or standards to estimate accuracy, effectiveness, economic benefits, etc.

Create
Be able to put parts or elements together in such a way as to show a pattern or structure not clearly there before; be able to identify which data or information from a complex set is appropriate to examine further or from which supported conclusions can be drawn.

The following pages provide examples of cognitive levels that item-writers can use to develop more complex types of items.

Cognitive Levels And Examples

What do you want candidates to do? And how do you want them to do it?

REMEMBER a fact (Knowledge)

Who makes the final determination regarding the distribution of the audit report?

(A) The lead auditor
(B) The audit group manager
(C) The auditee
(D) The client

UNDERSTAND (comprehend) symbols/directions

Producer's risk is defined as the probability of

(A) rejecting a good lot
(B) rejecting a bad lot
(C) accepting a good lot
(D) accepting a bad lot

APPLY abstracts to a concrete situation

Purchasing orders are to be checked for correctness during the audit of a purchasing department. Which of the following sampling plans would be most appropriate?

(A) Review 10% of all purchase orders.
(B) Randomly sample an amount of orders processed on the day of the audit.
(C) Review the first 10 orders processed on the day of the audit.
(D) Use MIL-STD 105 to establish a sampling plan.

ANALYZE data or information

Cases containing 24 bottles of antibiotics are shipped to drug stores. A pharmacist has doubts about the potency of the drugs and decides to have five bottles from a case tested. What is the probability that none of the tested bottles will prove to be defective if in fact there are 10 defective bottles in the case?

(A) 0.0010
(B) 0.0059
(C) 0.0471
(D) 0.2355

EVALUATE a situation

A lot composed of 500 units is submitted by a supplier whose past history indicates that about 1% defectives should be expected. A random sample of 30 units is collected from the lot. Which of the following probability distributions could be best used to make predictions about the lot?

(A) Normal
(B) Weibull
(C) Poisson
(D) Exponential

CREATE (synthesize) parts to create a whole

A supplier evaluates all characteristics on a sample of 75 components that is taken twice a day. Each component is then recorded as either conforming or nonconforming. What type of control chart would be used for evaluating this data?

(A) MR chart
(B) c chart
(C) u chart
(D) p chart

Creative Writing for Tests

There are three little-known facts that you need to be aware of before you begin writing items.

- **Most people think all tests contain at least some trick questions, and this is where they will learn how to write them.**

  [If I don't tell you this, you'll find it out the hard way by writing a trick question without knowing it.]

- **Written-English is not our first language.**

  [We have to translate everything from thought-English and spoken-English to written-English, and very often we are not good translators.]

- **You have a little edit switch inside your head that you have to turn to "off" before you can write effectively.**

  [Now is when you turn it off. I'll tell you when to turn it back on!]

Some Trick Questions

See if you can figure out what makes them tricky.

1. What is the best way to prevent emergencies from happening?

(A) Don’t get overwhelmed
(B) Plan ahead
(C) Always travel in pairs
(D) Call 9-1-1

2. Which of the following is considered “brain food”?

(A) Science fiction
(B) Non-fiction
(C) Tuna fish
(D) Meditation

3. What is the best way to travel to NYC?

(A) Fly
(B) Drive
(C) Subway
(D) Bus

4. Which of the following is NOT considered a fruit or vegetable?

(A) Pasta
(B) Potatoes
(C) Yams
(D) Maize

5. Which of the following countries, if any, built the Statue of Liberty and sold it to the United States?

(A) Britain
(B) France
(C) Germany
(D) None of the above

Writing Good Items Starts With Brainstorming

1. Identify a topic to focus on.

2. Brainstorm for three to five minutes in your small group: Think of all the words and phrases that are connected to that topic. Also, think of any work-related incident or story that has to do with the topic.

3. Step back from the flipchart and look over the words, then go around the table and ask each person to add to the list or pass. When everybody has passed twice in a row, go on to step 4.

4. Craft broken sentences for stems.

5. Loosely corral brainstormed notions for options. (A corral will hold anywhere from three to 10 untamed notions.)

6. Reshape the notions as broken sentences for stems.

7. Grab another bunch of loose notions and corral them with the new stems.

8. Begin honing the stems and options into raw items by mending the broken sentences and selecting four notions that get along well, only one of which is the key.

9. Repeat steps 4 through 8 until all the corrals are empty and all the broken sentences are mended into stems.

10. Ask each team member to recount his or her work-related incidents on this topic and capture the essence of them for further development as items.

ASQ ITEM FORM

BOK (or Content Code)
(to lowest level/subtopic):

Text Reference =
Author **Title** **Edition** **Year** **Page #:**

[ITEM]

Written by: _____
Reviewed by: _____ Key: _____

Reviewer's Notes:

Solution/Rationale

(A)

(B)

(C)

(D)

Using the Item Form

Fill in the BOK area being tested,
down to the
subtopic
level.

Provide specific information on the reference book:
author,
title,
edition,
page number (don't forget the page number!!!)

Fill in the
"Written by _____" line.

Be sure to write all four options
and label them A, B, C, D (not 1, 2, 3, 4).

Suggestions for Writing Good Items

Brainstorm items with your team. Capture "branch" ideas as you work together on one item or topic and save them to generate items in other areas of the BOK.

DO	DON'T
Item as a whole	
♦ Test knowledge of important facts and concepts	♦ **Do not** copy questions or sentences from *any* resource material
♦ Test one *main* idea at a time	♦ **Do not** use primers or refresher course materials as references
♦ Be as succinct (brief) as possible, but… →	♦ **Don't** be brief to the point of obscurity!
Stem	
♦ Create a clear and concise stem	
♦ Specify the complete task for each item	
♦ Use positive-stem formats	
♦ Keep the stem as short as possible to avoid wordiness, but … →	**Don't** be brief to the point of obscurity!
Options	
♦ Write only one correct answer	♦ **Don't** use specific determiners such as "always" or "all" or "never"
♦ Develop distractors that would be plausible even to uninformed test takers	♦ **Don't** match words between the stem and the key
♦ Write distractors that test common misconceptions	
♦ Keep options approximately the same length for each item	♦ **Don't** make the key the longest and most detailed.

A Review

Points to Remember When Writing Multiple-Choice Items

1. Focus on concepts that are important for the candidate to know or understand.

2. Don't test trivia or nits (as in nit-picking).

3. Use language that is simple, direct, and free of ambiguity. The verbal difficulty of an item should be kept to a minimum. An item should not be a test of reading ability.

4. Double negatives are not allowed. If the stem asks the candidate to answer by identifying an option that is *false* or *incorrect* (negative stem), the options should be stated in positive terms.

5. When multiple items are based on a single situation or common stimulus material such as a graph or chart, make sure that each item is independent of the others. The candidate is expected to arrive at an answer from the information provided in the stimulus material, not from having correctly answered another question in the set. This is also known as Double Jeopardy: If I get No. 13 wrong, and No. 14 is based on that answer, then I will automatically get No. 14 wrong, too.

6. Make the distractors reasonable and have them include misconceptions or errors in technique that are typical in the industry. Certain rules of thumb, procedures, or processes that do not apply to the stated problem often provide good distractors.

HAVE FUN!!!

The following pages contain examples of how NOT to write test questions. They are 'optional' for those of you whose eyes aren't glazed over yet! Not required reading, but interesting in its own way!

Common Pitfalls in Question Writing

- **The Nondirected Stem, also known as items created by and for telepaths**

Usually an open-ended question that isn't *quite* asked…To see if you have developed an indirect stem, cover up the options. Would you still know what answer to look for in the options?

Sample: Redundancy:

(A) provides more that one means for accomplishing a given function
(B) is always needed in complex systems
(C) is represented mathematically by the sum of the probabilities of failures of each path
(D) is usually wasteful and unnecessary

- **The Nonspecific Stem**

This is a closed-stem question and is a complete statement, but it still doesn't provide sufficient information to enable the candidate to answer the question without reading the options first.

Sample: Which of the following elements is important?

(A) Operation schedule
(B) Scheduled maintenance time
(C) Environment
(D) Cost of the system

The most obvious fall out of these items is that the answers are all over the place. Cure: brainstorm with the team; identify one key issue to test; discuss common misconceptions.

- **Never say Never; Always avoid Always**

Words such as **all, always,** or **never** are clues that the option is probably too specific to be correct.

Sample: Which of the following is the advantage of using the trace-forward method of auditing?

(A) It always shows the processing flow throughout the company.
(B) It is only used as a method for training auditors.
(C) It is never practical for all types of audits.
(D) It identifies deficiencies at the front end.

- **The Frustrated Novelist's Stem**

Items should be as short as possible while conveying essential information. There should be no teaching in the stem.

Sample: After a branch or division quality systems audit, an "exit briefing" of concerned personnel is conducted by the audit team, and a variety of staff or personnel from the audit team and the organization being audited meet, in person, to discuss the audit results. Which of the following is usually a violation of quality audit integrity?

(A) Identifying major quality system deficiencies found during the last audit
(B) Comparing audit results with the branch/division last audited
(C) Obtaining corrective action commitments
(D) Highlighting key areas of improvement since the last audit

Symptom: Your eyes glaze over. Cure: find the correct answer; identify key pieces in stem as needed; delete unnecessary information; and refocus the distractors.

- **Word Matching**

An option that uses the same key word or phrase as found in the stem is likely to be the key.

Sample: Reliability prediction is the process of

(A) estimating performance at a specific point in time
(B) predicting performance for a stated period of time
(C) telling "how to get there from here"
(D) defining operational safety requirements for equipment or system

- **The "Sore Thumb" Key**

In an attempt to make the key *REALLY* correct, an inexperienced item writer will often make it more specific than other options, thereby longer than others.

Sample: The primary objective of a pre-audit conference is to

(A) discuss schedules
(B) determine the scope of the audit
(C) introduce the audit team
(D) establish rapport and set ground rules for conducting the audit in a timely and efficient manner

- **Opposites Attract – or – The Total Universe in Just Two Options**

When two options are opposite of one another, one of them is usually the key.

Sample: A process safety review should cover

(A) any processes conducted on site
(B) any processes conducted off site
(C) those processes that have inherent risk for personal injury
(D) those processes that are provided as a service to customers

- **The Universe Includes the Solar System (also known as "Subsuming")**

When one option includes the same information as another option plus some, the more-is-better rule kicks in, and the candidate chooses the more inclusive answer for the wrong reasons.

Sample: Which of the following should be included in the evaluation of a supplier's capability?

(A) adequacy
(B) implementation
(C) adequacy and implementation
(D) adequacy, implementation, and documentation

Cure: It can sometimes be fixed by using "only," but a better solution is to find a single, best choice.

- **Apples and Oranges (also known as fruit, vegetables, and fork lift)**

Sample: When performing electrical stress screening, one of the aims is to

(A) see that the parts can withstand power surges
(B) cause mortalities to occur
(C) test the receiving inspection function
(D) measure ohms and amps

Cure: Identify the 'forklift' (i.e., the most obvious throw-away option) and replace it. Often it works to present a pair of vegetables and a pair of fruits, provided only one of them is keyable.