

Evaluation of the Present Level 2 Assessment System

July 14, 2004

Beverly Dianne Calhoun

IT 5990: Assessment and Technology, Summer Semester

Brent Wilson and Len Scrogan, Instructors

University of Colorado at Denver

Table of Contents

Table of Contents.....	2
Overview.....	3
Terms and Definitions	5
Organization Mission	6
Organization Objectives.....	6
Learning Outcomes	6
Evaluation Methods.....	7
Implementation of Assessment	8
Results.....	8
Decisions and Recommendations	8
Benefits for Using Simulation-like Questions.....	9
Examples: Before and After	10
Conclusion	13
Appendix A.....	14
References	15

Overview

Our organization uses a formative approach to assessments. The assessments measure the participant’s knowledge, skills, and ability to master the content of the course. Multiple-choice questions are the only type of question used for our assessments (this includes using graphics or long case scenarios). We constantly brainstorm ideas for new and creative ways to use multiple-choice questions. We want the participants to recall the learning better to use in their jobs and to prepare them for certification. Our Level 2 assessments are based on the second level of Donald Kirkpatrick’s four levels of evaluation.

- **Level 1:** Did participants react favorably to training?
- **Level 2:** Did participants learn the objectives of training?
- **Level 3:** Did participants apply their learning to change their behavior on the job?
- **Level 4:** Did the organization benefit from the training?

The following table shows the organization’s present Level 2 assessment considerations.

Topic	Category	Examples or Use
1	Use the following question types: <ul style="list-style-type: none"> ▪ Multiple-choice ▪ Multiple-choice using long case scenarios ▪ Multiple-choice using graphics 	Test knowledge, skills, and ability to master the content of the course for certification
2	Map questions to learning objectives and course content. Additional considerations: <ul style="list-style-type: none"> ▪ Verify 	Used to make sure the questions are pertinent to the learning.
3	<ul style="list-style-type: none"> ▪ Rules for questions: ▪ Include all information needed to answer the question ▪ Focus on one thought, problem, or idea ▪ Provide criteria when questions ask for judgment ▪ Questions should be independent of each other ▪ Avoid ambiguous questions ▪ Avoid absolutes ▪ Use terms common to all learners ▪ Make sure the question does not reveal the answer ▪ Randomize questions ▪ If using blanks, they should go at or near the end of the statement ▪ Use only one blank in a sentence ▪ Keep questions complete and options specific ▪ Questions should convey clear, complete, and precise information on how the learner is to respond to the question 	Used to create standards for how questions are created.

Topic	Category	Examples or Use
4	<p>Rules for distracters:</p> <ul style="list-style-type: none"> ▪ Four distracters <ul style="list-style-type: none"> ○ Mutually exclusive ○ Avoid absolute modifiers ○ Use terms common to all learners ○ Avoid clues that give away the right answer ○ All distracters should agree grammatically ○ Make distracters approximately the same length ○ Randomize answers 	<p>Wrong: When using the configurator to design an S8300 Media Server with G350 Media Gateway, how many port circuit packs does the configurator select by default?</p> <ul style="list-style-type: none"> A. 16 B. 20 C. 22 D. It depends on the type of circuit pack. <p>Right: When using the configurator to design an S8300 Media Server with G350 Media Gateway, how many port circuit packs does the configurator select by default?</p> <ul style="list-style-type: none"> A. 16 B. 20 C. 22 D. 24
5	<p>Graphics</p> <ul style="list-style-type: none"> ▪ Use appropriate sized and quality graphics ▪ Use graphics that pertain to the content 	<p>Used to visually enhance the assessment and add clarity to questions.</p>

Table 1: Our organization's present Level 2 assessment considerations.

Terms and Definitions

1. **Validity** measures the accuracy of the inferences made from an assessment's results. It answers the question, "Are we measuring the knowledge, skill, or ability to master the learning that a candidate must possess to be truly qualified at our pre-established minimum level of competency?" In simple terms, it asks, "Are we assessing the right people?"
2. **Reliability** refers to the accuracy and consistency of assessment scores and pass/fail decisions based upon them. Reliability is "repeatability." It measures how close the same person would get to their first score the next time the assessment is taken, and takes into consideration that no further studying has taken place. It answers the questions, "How accurate and consistent is this score?" "How much does this score vary from what this person should have scored were this assessment perfectly reliable?" In general terms, it asks, "How consistently are we assessing the right people?"
3. **Simulation-like-questions** are short-stem case scenario questions that provide learners with practice on the types of decision-making they will face in their real-world situations. These questions produce more powerful learning benefits than most other forms of questioning. They also have significantly fewer resources than full-blown simulations.
4. **Full-blown simulation questions** are long case scenario questions that provide learners with practice on the types of decision-making they will face in their real-world situations. Typically, these questions include a long scenario that has several questions associated with the scenario. Participant's must download the scenario and then answer the questions based on the scenario.
5. **Assessments** are any systematic method of obtaining evidence from posing questions to draw inferences about the knowledge, skills, attitudes, and other characteristics of people for a specific purpose.
6. **Formative** assessments that have a primary objective of providing practice for search and retrieval from memory for a participant and that provide prescriptive feedback (item, topic and/or assessment level).
7. **Summative** assessments are usually quantitative and the primary purpose is to give a definitive grade and/or make a judgment about the participant's achievement. If this judgment verifies that the participant has met an established standard indicative of special expertise, the judgment may confer "certification".

Organization Mission

The principal goal for our assessments is to create valid and reliable measurements of the participant's knowledge, skills, and ability to master the content of the course to prepare them for certification in their specific occupation or for use in their occupation.

Organization Objectives

In essence, a formative assessment approach examines knowledge, skill, and ability of the participant identified to master the content of the course. This approach blends well with the mission statement above and allows for the following objectives:

- Create valid, reliable assessment questions
- Measure the participant's knowledge, skills, and ability to master the content of the course
- Expand each participant's application of his/her knowledge, skills, or ability to master the content of the course

Learning Outcomes

The conclusions made from the assessment results should show that the assessment is consistently measuring the participant's ability to master the course content. The following are the learning outcomes the participants should demonstrate to show they have mastered the course content:

- Improved Knowledge – “What does the participant know?”
- Improved Skill – “What can participants do?”
- Improved Ability to Master the Course Content – “What has the participant learned?”

Evaluation Methods

The assessments are completed online and the scores are automatically entered into the participant's file. Reports are run to determine whether the participants have met the learning outcomes. The following table is an example of what a report looks like for an assessment question.

Assessment Item Classical Statistics Report										
Item Count	Item ID	Item Type*	Form	Tried	Right	P-Value	Point Biserial	90.0% Sig.	Reliability Index	Average Time
1	Q1	m3	beta	62	20	0.32	0.157	0.210	0.07	42.6
2	Q2	m4	beta	62	34	0.55	0.405	0.210	0.30	53.2
3	Q3	m2	beta	62	55	0.88	0.308	0.210	0.22	41.2
4	Q4	f	beta	62	14	0.23	-0.137	0.210	-0.06	47.3
5	Q5	f	beta	61	12	0.19	0.190	0.212	0.08	51.0
6	Q6	m5	beta	62	44	0.71	0.497	0.210	0.36	34.0
7	Q7	m1	beta	62	54	0.87	0.453	0.210	0.15	34.9
8	Q8	m1	beta	62	58	0.94	0.387	0.210	0.10	19.4
9	Q9	m1	beta	62	62	1.00	0.000	0.210	0.00	21.9
10	Q10	m1	beta	62	38	0.61	0.207	0.210	0.10	43.2

***Item Types:**
 tf = true/false, f = free response, m1 = multiple-choice, one option correct, m2 = multiple-choice, two options correct, m3 = multiple-choice, three options correct, m4 = four options correct, m5 = multiple-choice, five options correct, dd = drag and drop, pc = point and click, cg = complex graphical.

Definitions:
p-value – The proportion of candidates who answered correctly.

- 0.50 = 50% answered correctly
- 1.00 = 100% answered correctly
- 0.61 = 61% answered correctly

Point Biserial Correlation Coefficient – an indication of which candidates answered correctly.

- 1.000 = Perfect correlation (only knowledgeable candidates answered correctly)
- 0.500 = Knowledgeable candidates answered correctly in higher proportions than unknowledgeable candidates
- 0.000 = No correlation (knowledgeable and unknowledgeable candidates answered in equal proportions)
- -0.500 = Unknowledgeable candidates answered correctly in higher proportions than knowledgeable candidates
- -1.000 = Perfect inverse correlation (only unknowledgeable candidates answered correctly)

Table 2: Example table from the Developing Certification Test Items Guide (2003).

Implementation of Assessment

Our organization creates courseware to help prepare the participant to obtain the knowledge, and skills needed to become certified. The Level 2 assessments are how we measure the participant's ability to master the course content. The instructional designers create the courseware and the Level 2 assessments.

Our organization is constantly reviewing the assessment process for ways to improve the validity and reliability of questions for the assessments. Methods for creating questions are reviewed and when necessary, classes are conducted to teach the instructional designers how to write valid and reliable questions. Assessment standards and test scripts are used to help develop consistency between instructional designers in creating questions for assessments.

Results

Even with these processes in place, the results for the organization are not always good. Our organization's approach to creating formative questions is to use only a multiple-choice question format, with a recent trend in using multiple-choice questions that have graphics or long case scenarios. These scenario-based questions are working better; however, some assessment participants have had problems or issues with the following:

- Vague instructions for using the long case scenarios
- Misunderstanding the instructions for using the long case scenarios
- Ability to comprehend the long case scenarios and apply it to questions

These problems or issues may cause a participant not to pass the assessment. At the same time, the scenario-based questions have improved the validity of the learning experience. In addition, the use of graphics presents the question or scenario visually and helps represent more real-life work situations.

Decisions and Recommendations

Based on the positive results from the graphic and long case scenario questions, our organization has made a decision to use more of these types of questions. Because our organization is measuring the participant's knowledge, skills, and ability to master the content of the courseware, multiple-choice questions still need to be used; however, now the questions will be based more on real-life work situations and visual references.

One of the challenges to the long case scenarios is when participants have a hard time reading a long scenario and applying it to several questions. For this reason, I am making the following recommendations:

- **Use short-stem simulation-like questions**

Will Thalheimer, Ph.D. states simulation-like questions present the participant with realistic decisions (<http://www.work-learning.com/ma/publications.htm>, Writing Simulation-Like Questions, February 2003). By using short simulation-like questions, the confusion of downloading a long scenario and knowing where to use the scenario within the assessment would not be an issue. This type of question uses fewer resources than the longer case scenarios and produces a more powerful learning experience.

According to Thalheimer's research, writing these advanced types of questions put your instructional efforts high above the norm in terms of learning improvements. Retrieval practice

improves learning from 30% to 100%, and mirroring the performance context in your questions can improve learning from 10% to 55%. If feedback is provided (15% to 50%), repetitions (30% to 110%), and other instructional supports, learning outcomes significantly improve (<http://www.work-learning.com/ma/publications.htm>, Writing Simulation-Like Questions, February 2003).

- **Use appropriate feedback**

Feedback helps learners to reprocess the original question, which then doubles the time spent processing the learning point for a question. The goal is to use feedback to help strengthen the learner's retrieval of the correct learning points needed master the content of the course. Feedback is best used for the incorrect answers and is not necessarily beneficial for correct answers unless given as an optional choice for the learner to view if needed. Incorrect feedback reminds the learners of appropriate actions they take in real-life work situations and should be tailored to the answer. Giving detailed feedback only for incorrect answers emphasizes the learning point to the participant, which is not needed if they correctly answered the question. Thalheimer states, "Feedback on questions helps learners to overcome their misconceptions" (<http://www.work-learning.com/ma/publications.htm>, Writing Simulation-Like Questions, February 2003).

Benefits for Using Simulation-like Questions

According to Thalheimer, there are many benefits or arguments for advocating the use of simulation-like questions. The following are a few of the most important ideas I think would be beneficial for our organization (<http://www.work-learning.com/ma/publications.htm>, Writing Simulation-Like Questions, February 2003):

- Simulation-like questions produce similar benefits to those generated by full-blown simulations, but they do it at a fraction of the cost. Simulation-like questions do not require the costs of multimedia development. They do not incur costs for video and audio production, talent and talent recruitment, editing, studio rental, and project management associated with such efforts.
- Simulation-like questions utilize the same five learning factors that full-blown simulations do. They (1) provide realistic decisions set in realistic situations, (2) prompt learners to make decisions and retrieve information from memory, (3) provide realistic or didactic feedback, (4) use multiple scenarios covering the same learning points, and (5) space repetitions over time.
- Simulation-like questions can improve learning outcomes by 50 to 190%, the same rate possible through full-blown simulations.
- Simulation-like questions are easier to modify than multimedia-based simulations.
- The use of simulation-like questions is supported by proven research-based findings compiled from the world's best-juried journals on learning, memory, cognition, instruction, and performance.
- Learners seem to like simulation-like questions because they challenge them to perform in realistic situations. Learners feel they should know what to do, and hence they stay motivated throughout the learning.
- Because they are easy and inexpensive to create, simulation-like questions can be developed to repeat key learning points.

Examples: Before and After

A question checklist for creating short-stem case scenario question is in [Appendix A](#). Below are examples of the old and new format for assessment questions in our organization.

Old question format:

Learning Objective: Identify the proper placement of a network card when installed in a Windows server.

Question: What slot is a network card installed in a Windows server?

- A. One
- B. Two
- C. Three
- D. The one the customer wants it in

New question format:

Learning Objective: Identify the proper placement of a network card when installed in a Windows server.

Question: Your customer wants to connect all its computers together on one network. You have gone to the customer site and need to install a network card in their Windows server. There are several slots available and you need to decide which slot is the best one to install the card. Some slots are designated for other cards, so it is important that you install the card in the correct slot.

Question: Which slot do you install it in?

- A. One
- B. Two
- C. Three
- D. Four

Other possible multiple-choice question formats (These examples are from Les Scrogan’s (our teacher) presentation to our IT5990 class on June 22, 2004):

Multiple-choice Question Type	Example
1. Construct items in a form different from that originally presented.	Which of the following represents the meaning of the 4 in the problem 3)12? a. XXX XXX XXX XXX b. XXXXX XXXXX c. XXXXXXXXXXXXX d. XXX

Multiple-choice Question Type	Example
<p>2. Use novel pictorial materials to measure principles that require participants to apply knowledge.</p>	<p>The circle graph shows how one 6th grade class voted for class President. If there are 32 students in the class, how many did not vote at all?</p> <p>A circle graph divided into four segments. The largest segment, on the right, is labeled '1/2 Jenny'. The top-left segment is labeled '25% No'. The bottom-left segment is labeled '1/8 Suzy'. The bottom-right segment is labeled '1/8 Al'.</p> <p>A. 6 B. 8 C. 10 D. 12 E. 14</p>
<p>3. Provide for a condition contrary to fact.</p>	<p>If Robert Gray had not discovered the Columbia River in 1792, what country would have profited the most at that time?</p> <p>a. England b. France c. Russia d. Spain</p>
<p>4. Use analogies to measure relationships.</p>	<p>Governors (A) are to states (B) as presidents (C) are to</p> <p>a. their cabinets b. the electoral college c. the Senate d. the United States e. their vice-presidents</p>
<p>5. Select examples of principles or concepts.</p>	<p>“and, spite of pride, in erring reason’s sprite, One truth is clear: Whatever IS, is RIGHT.”</p> <p>This passage is a typical example of which of the following poetic styles?</p> <p>a. Classical b. Modern c. Romantic d. Victorian</p>
<p>6. Discover relationships among similar topics.</p>	<p>A assessment item for students who have studied different woodwind instruments could be:</p> <p>a. conical bodies b. reeds c. register keys d. tone holes</p>
<p>7. Identify assumptions and analyze criteria.</p>	<p>What criterion was used in the text to divide the colonial period in to New England, the South, the middle states, and the frontier?</p> <p>a. Economic b. Political c. Religious d. Social</p>


Multiple-choice Question Type	Example
<p>8. Use charts and tables.</p>	<p>Examine this chart:</p>  <p>This chart best demonstrates the concept of:</p> <ul style="list-style-type: none"> a. causation b. Malthus' theory c. efficiency d. inefficiency e. opportunity cost

Table 3: Other possible multiple-choice question formats.

Conclusion

Creating valid and reliable Level 2 assessments are imperative to an organization that uses questions to measure a participant's knowledge, skill, and ability to master the content of the course in order to be certified. It is the responsibility of all involved in the process of creating questions to make sure the end result, (the assessment) is an effective measurement and meets the participant's needs in becoming certified.

According to Thalheimer, the research shows that the learning benefit of questions can produce significant learning and performance benefits, potentially improving learning by 150% or more (<http://www.work-learning.com/ma/publications.htm>, Learning Benefits, January 2003).

Creating higher order thinking questions in the form of short-stem case scenarios allows the participants to have a more powerful learning experience. Participants should then be able to bring that learning into their real-life work situations. One of the most important benefits of short-stem case scenarios questions is that they give the participants practice in the retrieval process, which then allows for better learning retention of knowledge, skills, and ability needed master the content of the course. Short-stem case scenario questions are an effective way to provide and achieve learning gains that meet and exceed the organizations objectives for assessments.

Appendix A

No.	Question-Writing Checklist	Check if complete
1	Determine that simulation-like questions are appropriate, and understand how they will be used in your instruction, how they will benefit your learners, and how they will improve the performance of your learners when they return to their jobs.	
2	Determine overarching objectives. Why are you writing these questions? What do you want your learners to know? What situations will your learners face on the job? By answering these and other general questions for yourself, you will create some initial guidance for your writing. Make sure you write down your answers and keep track of your notes so that you will remember them.	
3	Learn the topic material. This is vital so that you can create rich, meaningful questions and so that you can determine how the material applies to the learners' on-the-job performance situations.	
4	Learn about the learners' performance situations. If you are going to create meaningful scenarios that mirror different performance situations, you have to know what these performance situations are like. Although you ought to have a general sense of the learners' day-to-day activities, it is most important to focus on the specific situations relevant to the topic.	
5	Create a list of all of the specific learning points you want learners to know. Because these are simulation-like questions, you are developing, try and keep most of your learning points focused on cause-and-effect relationships. In addition, keep in mind the following question as you develop your list of specific learning points: "What do you want your learners to do, and in what situations do you want your learners to do those things?" By keeping in mind the cause-and-effect relationship and the situated actions you want your learners to be able to perform, you will be able to create the most powerful types of questions.	
6	Begin writing a simulation-like question for each learning point.	
7	Write other questions that seem relevant, that focus on performance situations not covered, or that intrigue you. Do a reality check. Ask yourself whether the question really teaches a key learning point.	
8	Create brief feedback for each question or each decision choice. Feedback is most beneficial if it is very concise and direct. Praise for correct answers and empathetic responses for incorrect answers are not useful.	
9	If you plan to provide additional practice—generally a good idea—write a second or third question for each learning point. Where applicable, vary the decision choices. Where appropriate, vary the background situation described in the scenario. Consider varying the question type and providing learners with more difficult types of questions, even the more realistic recall questions.	

Appendix A: (Work-Learning Research by William Thalheimer, PhD at <http://www.work-learning.com/ma/publications.htm>, February 2003).

References

- Bresciani, M. (March 18, 2003). *An outline for assessment plans*. Retrieved June 20, 2004, from Assessment Website for NC State University: http://www.ncsu.edu/undergrad_affairs/assessment/files/evaluation/outline.htm
- Bresciani, M. (nd.). *Writing measurable and meaningful outcomes*. Retrieved June 21, 2004, from Assessment Website for NC State University: http://www.ncsu.edu/undergrad_affairs/assessment/files/evaluation/writingoutcomes.pdf
- Prometric, a part of Thomson Corporation. (2003). *Developing certification test items*. Prometric. Participant's Guide.
- Scogan, L., (June 22, 2004). *Higher order thinking question sampler: other possible multiple choice question formats*. Examples from Les Scrogan's (our teacher) presentation to our IT5990 class.
- Shepherd, E., Godwin, J. (May 10, 2004). *Assessments through the learning process*. Retrieved June 20, 2004, from Questionmark Corporation website: <http://www.questionmark.com/us/Learningcafe/>
- Thalheimer, W. (August 2002, updated February 2003). *Simulation-like questions: How and why to write them*. Retrieved June 20, 2004, from Work-Learning Research: <http://www.work-learning.com/ma/publications.htm>.
- Thalheimer, W. (January 2003). *The learning benefits of questions*. Retrieved June 20, 2004, from Work-Learning Research: <http://www.work-learning.com/ma/publications.htm>.