Knowledge check questions:
Best practices for use of this instructional strategy

David Lewis, PhD
National Science Foundation
College of Public Health, University of South Florida, United States of America
drlewis@nsf.gov

Trudian Trail, MEd
College of Public Health, University of South Florida, United States of America
ttrail@health.usf.edu

Sandhya Srinivasan, MEd., MPH
College of Public Health, University of South Florida, United States of America
ssriniva@health.usf.edu

Sang Joon Lee, PhD
College of Public Health, University of South Florida, United States of America
slee3@health.usf.edu

Samantha Lopez, MEd
College of Public Health, University of South Florida, United States of America
slopez@health.usf.edu

Abstract: The session considers the effectiveness of knowledge check questions, during a multimedia presentation. Classic instructional design literature argues for including these multiple choice questions in an e-learning lesson. However the empirical literature concerning self-tests with multiple choice questions is mixed, and has often demonstrated that this instructional strategy leads learners to negative learning outcomes. This session describes this conflict and best practices for overcoming the issues associated with immediate feedback.

Introduction

Knowledge check questions are a relatively new instructional strategy that has begun to be employed worldwide in e-learning, particularly in government and academia. These self-test questions are introduced into a lesson as an interactive component, with the expectation that this interactivity focuses the learner's attention on relevant content to clear any misconceptions, before progressing further into the lesson. While there is little empirical work concerning knowledge check questions in e-learning environments, there is a rich literature which documents the use of immediate feedback with paper-based multiple choice tests. This session intends to discuss this instructional strategy in the context of the literature and provide a series of evidence-based best practices that promote e-learning and improve performance. The session will focus on how the National Science Foundation (NSF) and the University of South Florida (USF) have implemented knowledge check questions in their e-learning environments.

Background

Although it can be argued that computer-based training (CBT) has existed for decades, the development of these materials began in earnest with the introduction of Apple's HyperCard® (Apple, 1987) and QuickTime Player (Apple, 1991). Following the introduction of these innovative products, instructional materials began to be widely distributed via CDROM in the early 1990s (Lewis, Moreno, & Large, 2009). Later, with the development of the World Wide Web, instructional designers began to deploy materials as web-based instruction (WBI) or Web based training (WBT) (Khan, 1997; Lewis et al., 2009). An explosion of computer-based training materials created the need for standardization, so in 1996, the U.S. Department of Defense (DoD) developed a department-wide strategy called the Advanced Distributed Learning (ADL) Initiative (Clinton library, 2009). In 1999, President Clinton followed on this action to sign executive order 13111 (OPM, 1999) that established the "President's Task Force on

Federal Training Technology" (OPM, 1999), with the express purpose of consolidating federal training efforts within a single learning technology standard, later called the Shareable Content Reference Model (SCORM) (Clinton library, 2006). SCORM is a "high level collection of specifications and standards" (ADL, 2009a). The SCORM model is important for it provides an underlying structure for all courses developed for U.S. government employees. Knowledge check questions are a form of assessment described in the U.S. Government's SCORM model (CMU/ADL, 2006). In the years since its development the SCORM standard has extended well beyond its intended governmental audience, and is now being used internationally by academia, business and corporate environments (SCORM, 2009). SCORM has even been incorporated as an output option for most multimedia authoring systems like Adobe Captivate, Adobe Presenter, Articulate Presenter, Techsmith Camtasia (Berking, 2009b; Adobe, 2009a; Adobe, 2009b; Articulate, 2009; Techsmith, 2009).

U.S. governmental agencies including the Department of Defense (DoD), the Department of Homeland Security (DHS), and several of the armed services (U.S. Army, U.S. Air Force, U.S. Navy, & U.S. Coast Guard) have all produced design guidelines for SCORM-based courses (Army, 2006; DHS, 2007; Navy, 2007; USAF, 2004; USCG, 2007). In each case, it seems that knowledge check questions are a suggested because of the SCORM specification (Army, 2006; DHS, 2007, USAF, 2004, CMU/ADL, 2006). Even though this is the case, there is little in the way of instructional design guidelines incorporated into SCORM (Chew & Hua, 2008). SCORM simply serves as a structure within in which developers can work, to promote learning. This set of specifications describes an underlying model or structure for the construction of computer-based training. Like HTML, it acts as a structure within which developers can distribute content. However this HTML analogy, is somewhat lacking because SCORM can actually alter the presentation sequence of content to the learner, and alter a learning plan for the learner (CMU/ADL, 2006).

Most of the above mentioned governmental publications consider knowledge check questions as an un-scored item which is not necessarily reported to the learning management system, and they are primarily used in sequencing the instruction (e.g. Navy, 2007). It is often suggested that these questions be placed within a narrated presentation and for these items to provide immediate feedback on the material recently covered. The use of "knowledge check questions" is an instructional strategy employed by instructional designers as they develop instructional sequences.

There is actually a good deal of classic instructional design literature which supports strategies like knowledge check questions, or self tests. Classic design literature like the writings of Gagné suggest that knowledge check questions reinforce previously covered material because they 1) will catch the attention of the learner and; 2) focus the learner on relevant material (Gagné, 1965). Merrill (1965) even developed instructional models that describe the inclusion of questions within a learning sequence. He used this correction/review process to help learners, those who initially missed material during instruction, were able to recover and then later perform as well as those who initially understood the material of during the presentation. Given the findings of this instructional technology literature review there is a good reason to believe the argument of suggesting knowledge check questions during e-learning courses.

The learning that results from this instructional strategy is due to the content presentation which precedes the knowledge check question, the question itself, and the presentation of feedback which follows. Even though learning does occur during the content presentation, this article is about the use of knowledge check questions and the issues associated with this instructional strategy. Therefore it will concentrate on the learning that occurs, as a result of the feedback presented during the question/feedback instructional sequence. Given the prevalence of SCORM and knowledge check questions as an instructional strategy, one would expect empirical findings to support the use of this instructional strategy. However the empirical literature on this subject is at best mixed (Mory, 2004), and often finds negative outcomes for those receiving immediate feedback (Brackbill, Bravos, & Starb, 1962; Kulik & Kulik, 1988; Kulhavy & Anderson, 1972).

Learning that results from "knowledge check questions"

The literature concerning this instructional strategy is quite old, since the instructional strategy of providing feedback following multiple choice questions has been explored extensively (Kulik & Kulik, 1988; Skinner, 1958). While a behaviorist paradigm is useful from a historic perspective, the literature of this review concentrates on more recent advances since this time, and concerned more with how feedback affects learning today in e-learning environments rather than programmed instruction. So this section considers feedback from a cognitive perspective, and more specifically considers two types of feedback: immediate and delayed feedback. As it turns out, the empirical literature over several decades has shown that immediate feedback following the question may be detrimental to learning (Brackbill et al., 1962; Kulik & Kulik, 1988; Kulhavy & Anderson, 1972). These studies and

many others have found that delayed feedback is more efficient and effective than immediate feedback (Mory, 2004).

Kulik and Kulik (1988) conducted a meta-analysis to consider the combined results of 53 studies which had considered the timing of feedback. They found that if feedback is supplied later than usual, many seconds to days later, learners perform significantly better on post-tests, than those who received immediate feedback. Kulik and Kulik results confirmed those of previous researchers (e.g. Kulhavy, 1977) and find that delayed feedback is a better instructional intervention for those learners that are trying to acquire verbal information provided in a multiple choice test, especially if that knowledge is tested days or weeks later.

Kluger and DeNisi (1996) mention that there is a widely held belief that feedback intervention improve performance. However, as they argued, responses to feedback are widely variable, and often produce negative or debilitating effects on performance. Unfortunately this negative influence on performance effect is widely ignored and poorly understood. According to Kulhavy and Anderson (1972) the history of this misconception ("the feedback hypothesis," the expectation of positive results given feedback), goes back many decades to Thorndike and his "law of effect" (Thorndike, 1913). As Kulhavy and Anderson (1972) discussed animal studies, like those conducted by Thorndike, are inconsistent with human studies, when immediate feedback is being considered. This is because humans react differently when presented with language-based materials (Brackbill et al., 1962; Kulhavy & Anderson, 1972; Sassenrath & Yonge, 1968). Humans use their language abilities to recall and relate their instructional materials to their prior learning (Sassenrath & Yonge, 1968). Feedback in this situation is primarily useful for the correction of errors (Kulhavy & Anderson, 1972; Guthrie, 1971) and therefore delayed retention is more beneficial to humans when they study language based learning materials (Brackbill et al., 1962; Kulhavy & Anderson, 1972).

Brackbill et al. were the first to document the differences in immediate and delayed feedback on performance (Brackbill et al., 1962; Kulhavy & Anderson, 1972). This is described as the delayed retention effect (DRE), but offered little explanation as to why humans react differently than animals during feedback interventions. Kulhavy and Anderson (1972) also considered this delayed retention effect (DRE). They found a positive learning effect for delaying feedback (a day or more). Unlike Brackbill et al. (1962), Kulhavy and Anderson proposed a theoretical explanation for the delayed retention effect. Both groups had provided learners with paper-based multiple choice tests, and then either given them the answers to the tests, either with the tests, or a day later. Both groups of researchers found students performed better on post tests, if they were provided the answers in a delayed manner. Kulhavy and Anderson proposed that this delayed response effect was due to the fact that learners would probably forget their incorrect responses, at a delayed presentation, thus having less interference from these responses. Kulhavy and Anderson (1972) described this as the "interference-preservation hypothesis" (p. 506) or what is now known as "interference-preservation theory" (Kulik & Kulik, 1988, p.80). According to this theory, an incorrect response may be too close in time, to the presentation of feedback. So a learner's memory of the incorrect response creates interference to the presentation of corrective feedback, causing the learner to fail to discriminate between the correct and incorrect response. During the delayed retention test, these learners are unable to recall the correct response, thus the error response is preserved (Chaparro, 1990).

Kulhavy and Stock (1989) proposed a model that shed some light on the mechanisms behind learner responses during instructional sequences. They proposed that much of the research prior to the late 1980s was based on an overly simplistic model of actual learner responses. This oversimplification is based upon a behaviorist "black box" conception of feedback and reinforcement. Kulhavy and Stock's new model "the certitude model of feedback" goes beyond this simple stimulus response approach, to consider the learner's intentions. Mory (2004) describes this model as the most comprehensive to date. Until this model, researchers had only considered a learner's response to be either correct, or incorrect. They state that a learner response is more complex and may be due to a number of factors. For instance, a learner could blindly guess the correct answer or choose an incorrect response because they misunderstood the question.

The Kulhavy and Stock's certitude model of feedback goes beyond learner feedback, to integrate the factors of learner confidence, feedback complexity, and error correction (Mory, 2004). This certitude model considers the certainty of the learner's answer (response certitude). Kulhavy and Stock measured how certain learners were, separately from the answer correctness. They explained that with a lack of certainty, there is some discrepancy in what the-learner knows and what they answer. Response certainty relies on a learner's metacognitive abilities (Mory, 2004) and when there is uncertainty a discrepancy exists. Kulhavy and Stock proposed that the greater a student's uncertainty the more time they will consider the feedback provided. They were able to demonstrate that the likelihood of a correct posttest response increased, with an initial response certainty level. In other words the more certain students were, the more likely they were correct. These researchers were even able to relate response times to certainty.

In the certitude model, feedback is considered to have two components, verification and elaboration (Kulhavy & Stock, 1989). Verification allows the learner to determine if their response is correct or incorrect. Any other information acts as elaboration, to provide an explanation for why the answer is correct or incorrect. Kulhavy and Stock use the term “elaborative feedback” to describe this additional information. Elaborative information alters the type and complexity of the feedback that the learner receives.

Knowledge check questions as an instructional design strategy

At the University of South Florida, instructional designers have employed knowledge check questions since 2004, in both graduate and undergraduate online courses. Knowledge check questions may be deployed in several ways: as questions with no feedback (not suggested), questions with feedback, and with or without scoring. NSF’s use of knowledge check questions is perhaps the most common use of this strategy. Like many governmental agencies, the National Science Foundation develops SCORM-based course presentations that include un-scored questions in the content presentation.

As a part of our course development process, both the University of South Florida and the National Science Foundation instructional designers work with subject matter experts to generate learning objectives for each lesson. Knowledge check questions are then developed from these objectives, and later strategically placed within narrated presentations, with intentions of aligning these questions with final assessments.

Typically, the NSF developers do not score these questions, so they only serve as practice and just-in-time feedback to students. This methodology has its advantages. At the University of South Florida we have taken it a step further, in that we have employed an innovative approach to using this instructional strategy. Questions are scored, not for grading purposes but to provide feedback to instructors and designers on how the lesson was received by students. In this case, the University of South Florida uses question scores as an indication of the effectiveness of the instructional materials. Scores help us develop a baseline for a lesson's effectiveness and can identify gaps in a lesson's ability to portray content. As feedback is received, designers are able to redevelop and refine the presentation based upon a student's inability to answer the knowledge check questions provided. This iterative approach allows for "continuous course improvement" and keeps both the instructor and instructional designer aware of how the lesson is being interpreted, assessed and finally accepted by students.

Finally, it should be stated that there is no perfect form of instruction or assessment. There are several disadvantages associated with multiple choice tests, both as a form of instruction and as a means of assessment. Dufresne, Leonard, and Gerace (2002) propose that this form of testing may be a false indication of conceptual understanding. This is because there is always some chance a student can guess the correct answer (Kuechler & Simkin, 2010). In addition multiple choice tests can be written at varying levels of difficulty. However these disadvantages are also offset by their advantages. An important advantage discussed in this paper is that multiple choice tests (or knowledge check questions with feedback) can act as a means of tutoring learners, to help guide and focus their attention during a content presentation. Even though other forms of assessment may test higher levels of thinking, they are not as efficient at assessing large quantities of factual or conceptual material. Certainly this form of questioning has become quite popular among those testing undergraduate students, because the grading of these exams can be automated (Dufresne, Leonard & Gerace, 2002). The point of this discussion is not to support any one form of assessment over another, just to state that each form of assessment has its own advantages and disadvantages. While many authors suggest teaching at higher levels of thinking, educators must at some point cover the basics.

Discussion and Conclusions

This literature review has attempted to summarize the literature as it relates to knowledge check questions and feedback. It provided a context for which knowledge check questions are used worldwide in both government and universities. It also described the literature of question feedback. In doing so it describes models of learning and how they propose feedback can either support or hinder learning while using multiple choice questions. Finally, this literature review attempted to provide some evidence-based guidelines as it explained the literature.

In short this review described two main points as knowledge check questions relate to feedback. It is suggested that when it comes to allowing a student to review feedback, that they 1) should be told if their answers are correct or incorrect (verification) 2) elaborative information can provide guidance for the learner; and perhaps most importantly 3) the timing of this feedback should be delayed if possible. If an online student is allowed to review a presentation more than once then this scenario is conducive to learning, for it provides both immediate and delayed feedback.

According to Kulhavy and Anderson (1972) delayed feedback allows a learner to forget incorrect responses in an effort to learn verbal information. Therefore it is not suggested that learners only be exposed to knowledge check questions once, or be provided with immediate feedback, and not allowed students to review the materials more than once. Kulhavy and Stock (1988) suggested the use of elaborative information in addition to the straight verification of the sample corrective feedback. This could potentially overload a learner and provide them with more information than they could retain in that time allowed.

References

ADL (2009a). *What is SCORM?* retrieved December 1, 2009 from http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/What%20Is%20SCORM.aspx

Adobe (2009a). *Adobe Captivate 4.0* [Computer program] San Jose, CA

Adobe (2009b). *Adobe Presenter 7.0* [Computer program] San Jose, CA

Apple (1988). *HyperCard Script Language Guide: The HyperTalk Language*. Reading, MA: Addison-Wesley

Apple (1991). *QuickTime 1.0* [Computer program] Cupertino, CA: Apple.

Army (2006). *United Stated Army Fort Knox Pam 350-70 Interactive Multimedia Instruction (IMI) Design and Development Guide*. retrieved November 22, 2009 from http://www.knox.army.mil/garrison/dhr/asd/pams/P350-70.pdf

Articulate (2009). *Articulate Presenter '09* [Computer program] New York, NY

Berking, P. (2009). *SCORM and authoring tools. Enabling interoperable object-based approaches to developing and presenting e-learning.* retrieved December 1, 2009 from http://www.2elearning.com/www/news/top-stories/single-news-article/article/scorm-and-authoring-tools-1.html

Brackbill, Y., Bravos, A., & Starb, R. H. (1962). Delayimproved retention of a difficult task. *Journal of Comparative and Physiological Psychology*, 55 (6) 947-952.

Chaparro , B.S. (1990). *An evaluation of STAR: A computerized tutorial in general psychology*. Unpublished doctoral dissertation, Texas Tech University, Lubbock, TX. retrieved February 28, 2009 from http://etd.lib.ttu.edu/theses/available/etd-02262009-31295005979843/

Chew, L. K., & Hua, T. G. (2008). Instructional strategies and limitations of the SCORM 2004 specification. *In the proceedings of the 16th international conference on computers in education (ICCE 2008)* (pp. 153-160) Taipei, Taiwan October 27-31, 2008 retrieved December 1, 2009 from http://www.apsce.net/icce2008/Workshop_Proceedings/Workshop_Proceedings_0153-160.pdf

Clinton library (2006). *Inventory for FOIA Request 2006-0175-F: Sharable Content/Courseware Reference Model (SCORM), Executive Order 13111, and The President's Task Force on Training Technology*. Retrieved November 30, 2009 from http://www.clintonlibrary.gov/Documents/FOIAS/more/2006-0175-F.pdf

Carnegie Mellon University/Advanced Distributed Learning (CMU/ADL) (2006). *The best practices guide for the design and development of SCORM assessments*. retrieved November 30, 2009 from http://www.adlnet.gov/About/Jointcolab/Documents/JADL/research/2005/assessments/learner_assessment_bpguide_final.pdf

Department of Homeland Security (DHS) (2007). Advanced Distributed Learning (ADL) Standards and Specifications Guide, Version 4.0. retrieved November 30, 2009 from http://portal.vertexsolutions.com/RTDC_1.8/rtdc_state/templates/DHS%20ADL%20Standards%20and%20Specifications%20Guide.pdf

Dufresne, R. J., Leonard, W. J., & Gerace, W. J. (2002). Making sense of students' answers to multiple-choice questions. *The Physics Teacher, 40*, 174-180.

Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart & Winston

Guthrie, J. T. (1971). Feedback and sentence learning. *Journal of Verbal Learning and Verbal Behavior, 10* (1) 23-28.

Khan, B. H. (1997). *Web-based instruction*. Englewood Cliffs, NJ: Educational technology publications, Inc.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Kuechler, W. L., & Simkin, M. G. (2010). Why is performance on multiple-choice tests and constructed-response tests not more closely related? Theory and an empirical Test. *Decision Sciences Journal of Innovative Education, 8* (1) 55-72

Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 68*(6) 506-512

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279 — 308.

Kulhavy, R. W. (1977). Feedback in Written Instruction. *Review of Educational Research, 47*(2) 211-232

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79–97.

Lewis, D., Moreno, M., & Large, J. (2009). Introductory videos: an analysis of student use patterns. *Journal of the Research Center for Educational Technology*, 5 (3) 68-79

Merrill, M.D. (1965). Correction and review on successive parts in learning a hierarchical task. *Journal of Educational Psychology, 56*(5) 226-234

Mory, E. H. (2004). Feedback research review. *In D. Jonassen (Ed.), Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Erlbaum Associates.

Navy (2007). *Navy ILE Instructional Content Style Guide Interactive Multimedia Instruction & Instructor-Led Training.* retrieved December 1, 2009 from https://www.netc.navy.mil/ILE/contentItems/Navy%20ILE%20Instructional%20Content%20Style%20Guide_20070815.pdf

Office of Personnel Management (OPM) (1999). *Executive order 13111: Using technology to improve training opportunities for federal government employees.* retrieved November 30, 2009 from http://www.opm.gov/pressrel/1999/eo.htm

Sassenhath, J. M., & Yonge, G. D. (1968). Delayed information feedback, feedback cues, retention set, and delayed retention. *Journal of Educational Psychology, 59* (2) 69-73.

SCORM (2009) *SCORM Explained*. retrieved November 30, 2009 from http://www.scorm.com/scorm-explained/

Skinner, B. F. (1958). *Teaching Machines*. *Science, 128*(3330) 969-977.

Techsmith (2009). *TechSmith Camtasia Studio 6.0*. [Computer program] Okemos, MI

Thorndike, E. L. (1913). *Educational psychology*. *Volume1: The original nature of man*. New York: Columbia University, Teachers College.

United States Air Force (USAF) (2004). *United States Air Force ADL Style Guide*. retrieved November 30, 2009 from https://adlhub.golearn.csd.disa.mil/dst/documents/ADL%20Technical%20Specifications_Final.doc

United States Coast Guard (USCG) (2007). *Standard operating procedures (SOP) for the coast guard's training system Volume 7*. retrieved November 30, 2009 from http://www.uscg.mil/hr/cg132/docs/SOP_7.pdf

Acknowledgement

This material was based on work supported by the National Science Foundation, while working at the Foundation. Any opinion, finding, and conclusions or recommendations expressed in this material; are those of the author and do not necessarily reflect the views of the National Science Foundation.