Journal of Educational Psychology
2011, Vol. 103, No. 2, 257–268

0022-0663/11/$12.00 DOI: 10.1037/a0022369

Varying Tutorial Modality and Interface Restriction to Maximize Transfer in a Complex Simulation Environment

Michael C. Mayrath
Harvard University

Priya K. Nihalani, Laura G. Torres, and Daniel H. Robinson
The University of Texas at Austin

In 2 experiments, 241 undergraduates with low domain knowledge viewed a tutorial on how to use Packet Tracer (PT), a computer-networking training simulation developed by the Cisco Networking Academy. Participants were then tested on retention of tutorial content and transfer using PT. Tutorial modality (text, narration, or narration plus text) was varied betweens subjects in both experiments, and simulation interface restriction (restricted or unrestricted) was varied between subjects only in Experiment 1. When PT's interface was unrestricted, students who received the narration tutorial performed better on the transfer task compared with students who received the text tutorial (statistically significant in Experiment 1 but not in Experiment 2). These findings extend the cognitive theory of multimedia learning (Mayer, 2005) by testing modality effects in new contexts and further specifying conditions of its applicability.

Keywords: cognitive load, modality, redundancy effect

Complex simulations are increasingly used in education because they allow students to apply and evaluate their knowledge, skills, and abilities through interactive tasks scored by embedded assessment systems. Embedded assessment systems score the student's performance on the basis of the actions and decisions made during the session (Gibson, Aldrich, & Prensky, 2006; Gredler, 1992, 2003). In the present study, we sought to identify instructional factors that would lead to optimal learning in a complex simulation environment by manipulating design features that may affect a learner's cognitive load.

Cognitive load theory states that humans have limited capacities to process information (Baddeley, 1986; Sweller, 1988; Sweller & Chandler, 1994). The theory is also central to the cognitive theory of multimedia learning (DeLeeuw & Mayer, 2008; Mayer, 2005). Cognitive load theory provides an effective framework for instructional design because it takes into consideration the limited attentional and cognitive resources of a learner, and the theory provides guidelines for instructional design by recommending that designers not overload a student with extraneous graphics or sounds. Over the past 20 years, Sweller and colleagues have developed a triarchic theory of cognitive load consisting of intrinsic, germane, and extraneous cognitive loads (Sweller, 2005; Sweller, van Merriënboer, & Paas, 1998).

Intrinsic cognitive load is the demand on cognitive resources caused by the complexity of selecting and organizing relevant sensory information (Sweller, 1988, 2005). First, there is an inherent difficulty in learning that is associated with how complex the learner finds the information presented. Second, germane cognitive load is the demand on resources required for integration of information between the working memory and long-term memory. Germane load requires meaningful mental effort, such as learners making sense of a concept, practicing skills, or understanding a mistake. Third, extraneous cognitive load is the demand on resources due to irrelevant stimuli, such as a poorly designed website that overloads the page with too much information or sounds. Extraneous load takes the learner's attention away from the task at hand and from the process of schema construction (van Merriënboer & Sweller, 2005). There is an obvious relationship between usability of a graphic–user interface and extraneous cognitive load.

Prior knowledge, or level of expertise, may influence cognitive load. Regarding intrinsic cognitive load, prior knowledge affects the internal complexity of the task relative to the learner (Sweller et al., 1998). Because intrinsic cognitive load is required for learning to take place, it is important that this load be managed so that it does not exceed the limits of working memory capacity (Kalyuga, 2007). For example, if intrinsic load is low for a particular task and learner, many cognitive resources are unused, and the task complexity (i.e., level of interacting task elements) can be increased (Pollock, Chandler, & Sweller, 2002). In contrast, however, if a learner has low prior knowledge within a specific domain, they experience high, unmanageable intrinsic load. The intrinsic load of learners with low prior knowledge can be managed by reducing the task complexity, dividing a task into smaller subtasks, and so forth. As learners acquire greater knowledge in the domain of interest, increasing task complexity can increase intrinsic load.

In complex simulations, expertise is a moderator of extraneous cognitive load (Kalyuga, Ayres, Chandler, & Sweller, 2003). For novices, a complex interface consisting of too many features may produce extraneous overload; however, for an expert, the simulation's interface may be too simplistic and could even result in an

This article was published Online First January 24, 2011.

Michael C. Mayrath, Graduate School of Education, Harvard University; Priya K. Nihalani, Laura G. Torres, and Daniel H. Robinson, Department of Educational Psychology, The University of Texas at Austin.

Correspondence concerning this article should be addressed to Michael C. Mayrath, 1112 Juniper Street, Austin, TX 78702. E-mail: michaelmayrath@yahoo.com

expertise reversal effect (Kalyuga, 2007). Building novice and expert versions of the same simulation is impractical, and designing two separate interfaces for the same simulation requires extensive programming. One solution is to simply restrict functionality of extraneous elements within a novice user's interface so that if novices have access only to the essential parts of the simulation needed for completing a task, then they will not waste valuable cognitive resources on extraneous simulation functionality. The restriction manipulation is related to another cognitive load effect, isolated–interacting elements effects (Pollock et al., 2002). The interacting element procedure occurs when a novice cannot process a multitude of elements within an instructional environment. The overwhelming elements in the environment overload working memory and, therefore, inhibit learning. In contrast, isolated elements can be managed and processed more effectively. Restricting part of a complex simulation's interface is a form of scaffolding similar to an isolated element procedure. It is also a relatively simple solution in terms of programming and, therefore, is cost effective. However, no research presently exists on what effect restricting access to parts of a simulation's interface has on the behavior and performance of a novice simulation user.

Packet Tracer: A Complex Simulation Environment

At the Cisco Networking Academy (NetAcad), a simulation called Packet Tracer (PT) was developed to be an experiential learning and assessment tool because it provides a high-fidelity, authentic experience and has an embedded assessment system that scores the student on the basis of performance in completing a task (Frezzo, Behrens, Mislevy, West, & DiCerbo, 2009). PT is used by thousands of Cisco students around the world to practice creating and troubleshooting networks (see www.cisco.com/web/learning).

Three pedagogical elements of PT's design create meaningful learning experiences for Cisco students (Frezzo & Stanley, 2005). First, PT has a simulation interface that allows students to experiment with, build, configure, and test networks, using virtual equipment and connections (Figure 1). Students are able to apply knowledge and skills by problem solving through drag-and-drop functions in the virtual work space. Second, PT has an advanced assessment system, the *assessment tree,* embedded in the program that allows instructors to create customized activities with automated assessments of the tasks. The assessment tree allows the instructor to check or uncheck whether specific task components will be assessed when comparing an initial model of the network to the answer model of the network (Figure 2). This built-in assessment provides a behavioral measure of the student's knowledge, skills, and abilities (Behrens, Frezzo, Mislevy, Kroopnick, & Wise, 2007). Third, PT provides authentic visual representations for all Cisco products (i.e., symbols, icons, switches, hubs, routers) and common computer-networking environments (i.e., computers, servers, building blueprints) that a NetAcad graduate needs to understand. Similarly, the virtual equipment looks and responds in the program as it does in real life. For example, PT has a virtual representation of the front of a PC. A user can turn the PC on or off by clicking on the power button. Finally, PT also provides an authentic representation of situations that networking engineers experience in real-world settings (Frezzo et al., 2009).

Given these three pedagogical elements of PT, it is reasonable to claim that the simulation facilitates an authentic and situated learning experience similar to the real-world context (Herrington

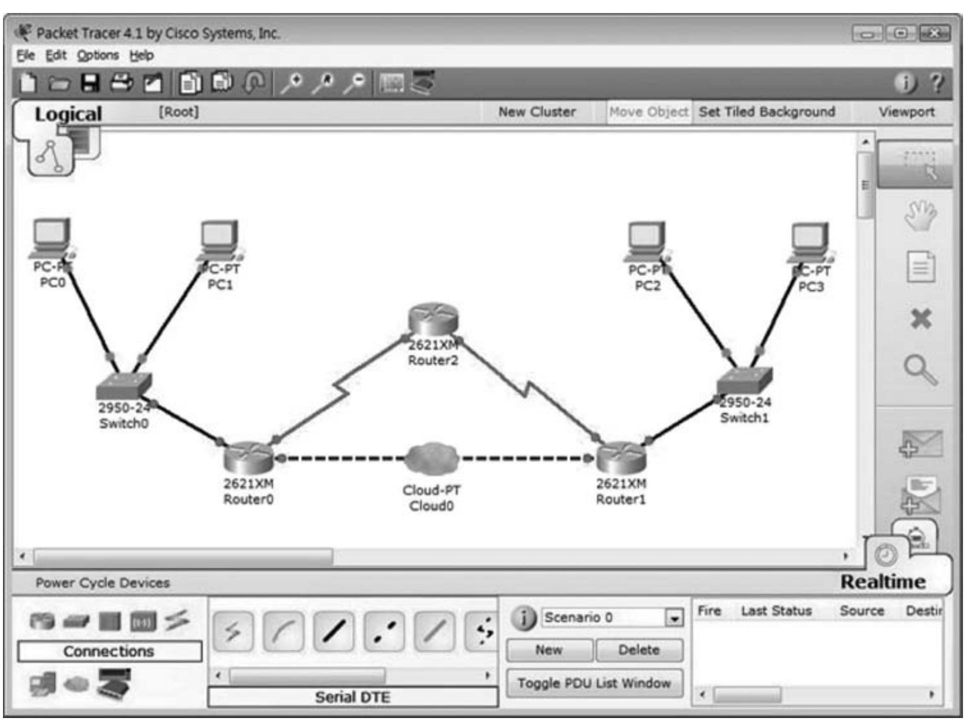

Figure 1. Packet Tracer's user interface. Packet Tracer (Version 5) copyright 2009 by Cisco Systems, Inc.

Figure 2. Packet Tracer's assessment tree. Packet Tracer (Version 5) copyright 2009 by Cisco Systems, Inc.

& Kervin, 2007; Lave & Wenger, 1991). At least one study has compared skills application of students who trained with PT to that of students who trained on real equipment and demonstrated that students who used PT for training performed on real equipment at a level equivalent to that of students who trained on real equipment (Frezzo et al., 2009; Nihalani, Mayrath, & Robinson, 2010). PT therefore provides researchers a solid platform for studying how simulations can be designed to increase a student's computer-networking knowledge, skills, and abilities while simultaneously assessing the student's performance. Currently, this simulation is used as a formative and summative assessment tool for all levels of Cisco students ranging from novice to expert. Because PT has expert functionality, novice students must first watch a series of animated tutorials called *My First PT Labs* prior to using PT. Additionally, PT is introduced to students in a face-to-face classroom setting so that they have support from a PT expert.

Simulation Challenges

There are two challenges associated with using complex simulations that may prove detrimental to the learner. First, tutorials are frequently needed before a student can utilize a simulation to its full pedagogical potential. A tutorial's effectiveness, however, is determined by how well it is designed, and numerous tutorials are poorly designed because developers are unfamiliar with design principles for multimedia learning (Clark & Mayer, 2003; Mayer, 2005). Second, experts and novices differ in their use of strategies to overcome limitations of working memory wherein only a few elements of information can be processed at any one time. Experts use *chunking strategies* to increase their processing capacity by integrating multiple elements of information (Bransford, Brown, & Cocking, 2000; Miller, 1956). Novices, in contrast, are unable to apply this same strategy if presented with the same number of information elements (Kalyuga, Ayres, Chandler, & Sweller, 2003). Cognitive load capacity also varies as a function of experts' and novices' varying availability of working memory resources. Thus, should complex simulations be designed differently for experts and novices? Or are there other methods that can be used to avoid overloading novice students' cognitive capacities while using an expert-level simulation interface? These two challenges are discussed next in greater detail.

Tutorials and Modality of Instruction

Simulations, such as PT, are complex and can easily frustrate the user when no training, or inadequate training, is provided on how to use the simulation. *My First PT Lab* tutorials are a series of 3- to 6-min animated Adobe Captivate screencasts that orient the user to PT's interface by showing how to drag network icons and symbols to the simulation's work space. The narration in the tutorial is delivered through pop-up text boxes that accompany the animation. There is no voice narration. Given recent research on how people learn, there may be more effective ways to present the information in these tutorials (Mayer, 2001).

According to components of the cognitive theory of multimedia learning (Mayer, 2001, 2005), specifically the *modality effect* and the *redundancy principle*, tutorials that simultaneously present an animation and a text-based narration cause students to split their attention by having to read the text narration while watching the animation. This split of visual attention has been shown to result in a decrease in the retention and transfer of knowledge. Research has shown that an animation accompanied by an audio narration is more effective on both retention and transfer measures compared with an animation with text boxes for narration (Ginns, 2005). Paivio's (1986) dual-code theory helps explain the split-attention effect, stating that humans process information through visual and auditory channels representing verbal and pictorial subsystems. Twenty years ago, Penny (1989) provided evidence that effectively utilizing both auditory and visual subsystems can increase working memory capacity because information is processed together rather than through only one channel.

Past research has shown that the *modalities* in which information is presented in multimedia learning environments can be a testable variable that informs both theory and practice. Modality research has focused on explanations of scientific processes, such as explanations of lighting, brakes, science, and botany (Mayer, 2001, 2005). However, there has been no research testing the modality effect in a tutorial that teaches novice students how to use a complex simulation. What is the best instructional delivery modality for such tutorials?

Simulation Interface and Cognitive Overload

Whether novice or expert in domain knowledge, it can take time to learn how to use a complex simulation's interface. In any specific domain, experts are at an advantage over novices because their working memory is more effectively utilized because of existing well-formed schemas (Kalyuga et al., 2003; Sweller, 1988). For example, if a person with no piloting experience had to land a plane, all the buttons and levers on the control panel would be overwhelming. What if the available options were reduced to only the controls needed to land the plane? This form of scaffold may minimize the options for error or distraction (Chang, Sung, & Chen, 2001). Similarly, restricting access to extraneous functions in a computer-networking simulation may reduce a novice's potential for error. Thus, a simulation should present novices with less information and fewer distracting stimuli; yet, simulations should also provide experts with advanced tools for hypothesis testing and exploration. How should developers design simulations that do not overload novices while also providing sufficient functionalities for experts?

Experiment 1

This study examines two distinct areas of research to address the challenges preventing students from utilizing the full pedagogical potential of simulations: (a) the design of tutorials that teach novices how to use a complex simulation and (b) the design of a complex simulation's user interface. More specifically, we aimed to answer the following questions.

How does instructional delivery modality (text only, voice only, voice plus text) of the *My First PT Lab* tutorial affect performance on dependent measures of retention and transfer? On the basis of prior research, we hypothesized that the high extraneous load conditions (voice plus text and text only) would overload students' visual channel, thereby resulting in poorer performance compared with students' performance in low-load conditions (voice only; Mayer, 2001, 2005). Thus, students' who receive the voice-only tutorial should outperform the other two tutorial conditions (Mayer & Johnson, 2008).

What effect does the restriction of PT's interface (restricted vs. unrestricted) have on these same dependent measures? There is no literature currently available that compares the effects of a simulation's interface being restricted or not. However, we expected that because these participants were novice PT users, those who received the restricted interface would experience less cognitive overload and therefore outperform those who received the unrestricted interface (Pollock et al., 2002). Further, is there an interaction between tutorial modality and interface restriction? On the basis of previous research and the present study's hypotheses regarding modality and interface restriction, it was predicted that there would be no interaction; instead, performance on the retention and transfer measures would indicate that modality and restriction would have an additive effect resulting in the best condition being voice only and restricted.

Thus, we sought to inform instructional designers of variables that affect student learning outcomes by first identifying optimal modalities for presenting a simulation tutorial and then examining the effect of restricting access to extraneous parts of a complex simulation's graphic–user interface.

Method

Participants and experimental design. We randomly selected 146 students from a subject pool at a large southwestern university. Students received course credit for participation. Because of a technological scoring issue (discussed further in the Materials and apparatus section), this sample size was reduced to 81. The sample was composed of 25 male and 56 female participants; in terms of grade classification, there were 45 seniors, 19 juniors, 11 sophomores, 3 freshmen, and 3 graduate students. Students were randomly assigned to one of six treatment groups in which tutorial modality (text, voice, or voice plus text) and simulation interface (restricted or unrestricted) were completely crossed in a 2×3 factorial design: text-only restricted ($n = 12$), text-only unrestricted ($n = 15$), voice-only restricted ($n = 14$), voice-only unrestricted ($n = 14$), voice-plus-text restricted ($n = 14$), and voice-plus-text unrestricted ($n = 12$) groups.

Materials and apparatus.

Tutorial. The 10-min tutorial was developed with Adobe Captivate to teach users how to navigate within PT's network environment and how to work with computer networks. The video included a 3-min introduction to fundamental computer-networking concepts and content from NetAcad's *My First PT Labs*. The videos are used by Cisco to walk students through the general features and functions of the PT simulation and also to teach networking students how to create, configure, and test a computer network. To maintain experimental control, students were not allowed to control the pace of the tutorial; thus, they were unable to stop, pause, rewind, or fast-forward. Figure 3 shows a screenshot of the tutorial.

PT. The PT simulation is both a training and an assessment tool. PT has a built-in authoring system that allows Cisco instructors to create their own performance assessments. The system first presents the instructor with an *answer network,* which is a fully functional virtual computer network and is represented by two components. The first component is the functioning network visualization interface where the user can experiment with virtual equipment that looks and responds in PT as it does in real life. For example, a computer's power button would be green to indicate that it is on rather than red to indicate that it is off. The second component is the assessment tree, which is essentially a checklist corresponding to the answer network. For example, the assessment tree would contain a line item specifying the computer's power button. Instructors then *break* the answer network, or the fully functional network, by unchecking items from the checklist. If turning on a computer were part of the student's task, an instructor would uncheck the line item referring to the power button (Figure 2). This broken network is referred to as the *initial network* and is the network that the students were tasked with fixing. Once students have completed the task, the program uses the assessment tree to compare the initial network (the task delivered to the student) to the answer network (the instructor's fully functional network) and scores students' performance.

Additionally, PT was designed with the ability to scaffold users by restricting their use of specific functions or capabilities of the simulation. Instructors may choose to restrict complex features of PT that are not necessary for novices to complete a specific task or they may restrict access to features, such as the help menu, during formal curriculum assessments. In this study, PT's interface was restricted for half of the students and unrestricted for the other half. Those features that were restricted, such as recording and playing back the user's actions, were not necessary for completing the transfer task. Figure 4 shows PT's restricted functions in red.

PT transfer test. Following the tutorial, students were given a PT-based transfer test. A computer network was already created in the PT work space; however, it was broken. Students were to troubleshoot, fix, and configure the broken network by applying part of the content shown in the tutorial. The time allotted was 10 min. A pop-up window appeared at the beginning of the activity with directions and a timer (also displayed in Figure 4). Additionally, the transfer test was automatically scored by PT's assessment wizard on a 100-point scale.

The first two stages of evidence-centered design—domain analysis and domain modeling—were used to build an assessment argument for the task (Mislevy, Steinberg, & Almond, 2003). The process included analyzing and mapping basic computer-

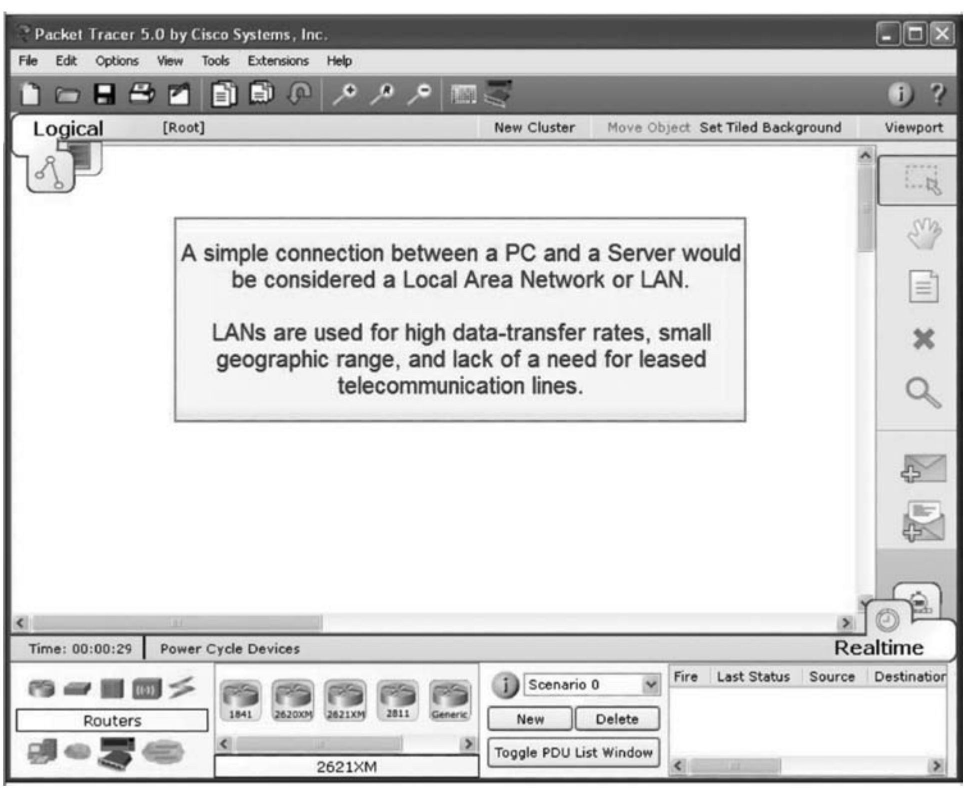

Figure 3. Screenshot of the text-only tutorial. Packet Tracer (Version 5) copyright 2009 by Cisco Systems, Inc.

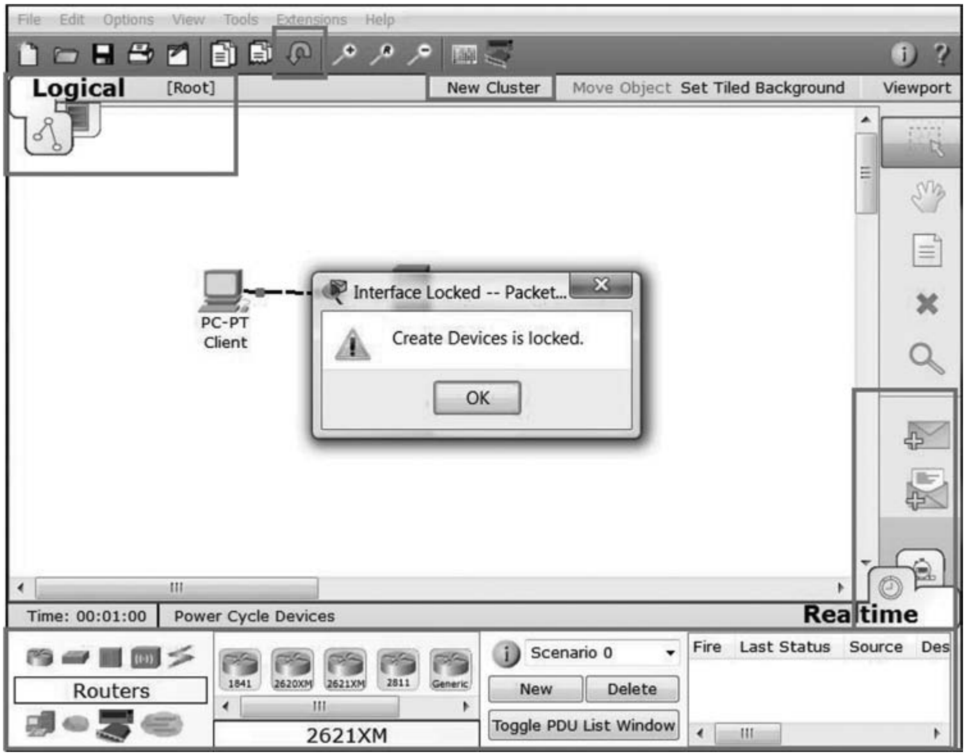

Figure 4. Packet Tracer restricted. Packet Tracer (Version 5) copyright 2009 by Cisco Systems, Inc.

networking knowledge, skills, and abilities related to Cisco's curriculum and then to observable variables produced while completing the PT activity. Applying evidence-centered design ensured that the instruction presented in the tutorial was matched to the tasks required for the PT activity and created a systematic procedure for establishing content validity.

Scoring issue. As described earlier, the PT assessment tree allows the author of an assessment to specify details that PT should check when comparing the initial network and the answer network; the assessment tree is very sensitive to changes between the two networks. A scoring issue occurred because students changed the name of either the computer or server, which caused the assessment tree to not recognize that device. When students clicked on the virtual computer, they were given many options regarding that specific computer. If the user changed the name of the computer or the server, two of the potential options, all points for that network component were lost. Thus, if the network problem was correctly diagnosed but the name of the computer was changed, the student would lose all points for the computer. This issue was not recognized until data collection began and required that data from the first 65 students be thrown out because there was no way of evaluating whether their scores accurately represented their performance. Directions during the remaining study sessions were modified to indicate that participants should not change the name of the computer or server.

Measures. Data collected were as follows: student demographics, a 10-item multiple-choice pretest of computer-networking prior knowledge, a 14-item multiple-choice retention measure that assessed declarative knowledge over content presented in the first half of the tutorial, the PT-based transfer test, and a self-report survey consisting of 12 items measuring cognitive load, three items measuring interest, and three items asking participants in the locked-down group about their response to the treatment. Except for the PT-based transfer test, all measures were delivered to students' computers through Survey Monkey, an online survey tool (see www.surveymonkey.com).

All cognitive load items were adapted from DeLeeuw and Mayer's (2008) research in which they used self-report items to find evidence of differentially tapped intrinsic, germane, and extraneous cognitive load. In other words, DeLeeuw and Mayer found that different measures of cognitive load were sensitive to different types of cognitive load. Intrinsic cognitive load was measured with four self-report items that asked participants to rate their mental effort during parts of the study. Two questions asked participants to rate their mental effort from extremely low to extremely high during the tutorial and while using PT. The next two questions asked participants to rate the complexity of the tutorial and PT on a 5-point scale (1 = *extremely simple*, 5 = *extremely complex*). Germane cognitive load was measured with three self-report items that asked participants to rate the difficulty of parts of the study on a 5-point scale (1 = *extremely easy*, 5 = *extremely difficult*). The items included how difficult it was to understand the tutorial, to remember information from the tutorial, and to use PT. Extraneous cognitive load was measured by four self-report items that asked participants to rate the usability of PT's graphic–user interface. The extraneous cognitive load items used a 5-point Likert scale (1 = *strongly disagree,* 5 = *strongly agree*) and included the following: "Packet Tracer was fun," "Packet

Tracer's interface (visual display) was easy to use," "Packet Tracer was frustrating," and "Packet Tracer's interface was overwhelming."

Procedure. The study was conducted in a computer lab. Each computer had an open folder on the desktop that contained the files for this study (i.e., the pretest, PT tutorial, PT transfer activity, and posttest). Within the folder, the files were labeled by their respective step for completion. For example, the file labeled *Step 1* was the Survey Monkey webpage for the pretest. Additionally, each computer was configured with an identification code that ensured anonymity and was also specific to one of the six treatment groups so that when students arrived at the lab, they were randomly assigned to a computer/treatment group. A member of the research team went through study protocol and also provided a paper-based copy of the four steps to be completed as well as additional information to successfully set up the network in the PT transfer activity.

Regardless of treatment group assignment, all students first completed the pretest (Step 1). They then chose the tutorial option that corresponded with their computer ID code and assigned treatment group (Step 2). Following completion of the tutorial, students were given 10 min to complete the transfer activity (Step 3) and finally the posttest and self-report items (Step 4). Students were instructed to not discuss the simulation, tutorial content, or activity with other students. Raw scores for the dependent variables (the retention posttest and PT transfer test) and the pretest of prior computer-networking knowledge were converted to proportion correct scores. Because 11 students' transfer scores were lost as a result of the PT activity scoring issue, the sample size was reduced from 81 to 70.

Results and Discussion

Table 1 displays the means, adjusted means, and standard deviations for the pretest, retention test, and PT activity transfer test scores for each of the six treatment groups. For all statistical tests, an alpha of .05 was used. A 2 (restricted vs. unrestricted) × 3 (text, voice, voice plus text) analysis of variance (ANOVA) conducted on the pretest scores indicated that the six treatment groups did not differ on prior knowledge of computer networking, $F(2, 75) = 1.71$, $MSE = 0.03$, $p > .05$, Cohen's $f = .21$ (small).

Students' pretest scores and retention scores were correlated ($r = .42$); thus, an analysis of covariance (ANCOVA) was conducted on the retention test with pretest as a covariate. There was no Modality × Restriction interaction, $F(2, 75) = 1.14$, $p > .05$, $MSE = 0.03$, Cohen's $f = .17$ (small), nor were there main effects for either modality, $F(2, 75) = 0.68$, $MSE = 0.03$, $p > .05$, Cohen's $f = .13$ (small), or restriction, $F(1, 75) = 0.15$, $MSE = 0.03$, $p > .05$, Cohen's $f = .04$ (small).

There was no statistical correlation between the pretest scores and PT activity transfer test scores likely because of the different constructs measured by the two tests. The pretest measured basic computer-networking terminology, whereas the transfer test measured procedural knowledge that was taught during the tutorial and applied during the PT activity. Thus, an ANOVA was conducted on the PT activity transfer test. There was a Modality × Restriction interaction, $F(2, 64) = 5.62$, $MSE = 0.03$, Cohen's $f = .40$ (large). Simple effects tests of modality within each restriction condition were conducted and revealed a modality effect within the unrestricted condition, $F(2, 31) = 3.51$, $MSE = 0.03$, Cohen's $f = .45$ (large). A Fisher's least significant difference (LSD) post hoc test indicated that students in the voice condition ($M = 0.84$, $SD = 0.08$) outperformed students in both the text-only ($M = 0.66$, $SD = 0.12$) and voice-plus-text conditions ($M = 0.70$, $SD = 0.19$). There was no modality effect within the restricted condition, $F(2, 31) = 2.70$, $MSE = 0.03$, Cohen's $f = .13$ (small).

We conducted 2 (restriction) × 3 (modality) ANOVAs on the self-report items. Additionally, because the items may also be viewed as ordinal measures, nonparametric Kruskal–Wallis *H* tests were conducted as follow-ups for any significant effects to control for Type I error inflation.

Significant effects were found for two cognitive load items; however, the other 10 items were not significant and, therefore, were not included in this discussion of results. A main effect of modality was found for the self-report item, "Packet Tracer's interface was easy," $F(2, 75) = 5.08$, $MSE = 1.16$, $p < .01$, Cohen's $f = .35$ (medium). Fisher's LSD post hoc tests revealed that both the voice ($M = 3.46$, $SD = 0.83$) and voice-plus-text ($M = 3.75$, $SD = 1.07$) groups perceived PT to be easier to use than did the text-only group ($M = 2.84$, $SD = 1.27$), $p < .05$. Thus, having voice in the tutorial positively affected how students perceived PT in terms of ease of use. The Kruskal–Wallis *H* test was consistent with the parametric results, $\chi^2(2, N = 81) = 8.70$, $p < .05$.

A main effect of modality was also found for the self-report item, "Packet Tracer was frustrating," $F(2, 75) = 4.06$, $MSE = 1.13$, $p < .05$, Cohen's $f = .32$ (medium). Fisher's LSD tests revealed that the text group ($M = 3.37$, $SD = 1.18$) found PT more frustrating than both the voice group ($M = 2.61$, $SD = 1.03$) and

Table 1
Means, Adjusted Means, and Standard Deviations for Proportion Correct Scores on Dependent Measures in Experiment 1

	Text only						Voice only						Voice plus text					
	Restricted			Unrestricted			Restricted			Unrestricted			Restricted			Unrestricted		
Measure	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
Prior knowledge	.53	.21	12	.49	.19	15	.59	.16	14	.56	.14	14	.46	.16	14	.59	.21	12
Retention	.61[a]	.22	12	.67[a]	.17	15	.58[a]	.11	14	.60[a]	.14	14	.64[a]	.17	14	.57[a]	.17	12
PT activity transfer	.72	.16	10	.66	.12	11	.64	.15	13	.84	.08	11	.78	.14	13	.70	.19	12

Note. The potential range of scores was 0–10 for the pretest, 0–14 for the retention test, and 0–9 for the PT activity. PT = Packet Tracer.
[a] Adjusted means using the pretest covariate.

the voice-plus-text group ($M = 2.73$, $SD = 0.96$), $p < .05$. Therefore, the text-based tutorial negatively affected how students perceived PT in terms of frustration compared with the voice groups. The Kruskal–Wallis H test was significant, $\chi^2(2, N = 81) = 6.90$, $p < .05$, confirming the parametric findings.

Two cognitive load self-report items were found to have statistical differences for the modality factor and helped to inform the Modality × Restriction interaction for the transfer test. The voice-only and voice-plus-text groups perceived PT to be easier to use and less frustrating than the text-only group. This is consistent with the finding where the voice-unrestricted group outperformed the text-unrestricted group on the PT activity. In terms of cognitive load, the text group reported having higher extraneous load (frustration) than the voice groups during the PT activity.

Because we found only modality effects when PT's interface was unrestricted, we decided to replicate Experiment 1, using the three unrestricted conditions with a larger sample to see if the modality effect would again emerge.

Experiment 2

The purpose of Experiment 2 was to replicate the unrestricted modality effect after correcting for design errors while increasing the sample size. Given both the findings from Experiment 1 and prior research on the modality effect, we hypothesized that the voice-only condition would demonstrate greater transfer than the other two conditions (Mayer, 2001). Further, we hypothesized that students within the text-only and voice-plus-text conditions would demonstrate greater levels of cognitive overload as evidenced by self-report data (Mayer & Johnson, 2008).

Method

Participants and procedure. We selected 160 students from the same subject pool as in Experiment 1. The sample was composed of 70 male and 90 female students and included 95 seniors, 37 juniors, 13 sophomores, 11 freshmen, and 4 graduate students. Students were randomly assigned to one of three conditions in a between-subjects design: the text-only unrestricted, voice-only unrestricted, and voice-plus-text unrestricted conditions. The materials and procedure were identical to those used in Experiment 1 for the unrestricted condition. We converted raw scores for the dependent variables (the retention posttest and PT transfer test) and the pretest of prior computer-networking knowledge to proportion correct scores, using the same method as in the first experiment.

Results and Discussion

Table 2 displays the means, adjusted means, and standard deviations on the pretest, retention test, and PT activity transfer test for each of the three treatment groups. Pretest scores correlated with both the retention and PT transfer test scores; thus, separate one-way ANCOVAs with modality (text, voice, voice plus text) as the between-subjects variable and pretest score as the covariate were conducted on both. Consistent with Experiment 1, no modality effect was found with the retention scores, $F(2, 156) = 0.059$, $MSE = 0.032$, $p = .94$, Cohen's $f = .03$ (small). Surprisingly, no modality effect was found with the PT activity transfer scores, $F(2, 156) = 0.429$, $MSE = 0.092$, $p = .65$, Cohen's $f = .07$ (small). A comparison of the pretest, retention, and transfer scores in Experiments 1 and 2 reveals that the students in Experiment 2 performed much better on the transfer test than those in Experiment 1. There were also much larger standard deviations for the scores in Experiment 2 than in Experiment 1. This warranted a closer examination.

Among the corrections we made to avoid losing data, as in Experiment 1, was a visible timer when students were completing the PT activity. For students who either ran out of time or finished with a score of 100%, an *end of activity* screen appeared. They were then supposed to save the activity to the desktop. If they did not save the file to the desktop correctly, they were directed to either close the file or restart the activity. Thus, several students restarted the activity and were allowed to replace their original score. We caught this after a few sessions and took care to explain this thoroughly in the directions given in the remaining sessions. Unfortunately, we were unable to separate those students who had restarted and completed the PT activity after a "practice" session from those who had completed the activity only once.

Because we have no reason to believe that any of the three conditions had more participants who restarted the activity than others, we decided to analyze the complete data set. However, we realize that this procedural error likely led to some measurement error (inflated scores for some students that led to increased group means and standard deviations) and also likely contributed to our lack of statistical power in replicating the statistical findings of Experiment 1. Thus, our findings are interpreted with caution.

Table 2
Means, Adjusted Means, and Standard Deviations for Proportion Correct Scores on Dependent Measures in Experiment 2

	Text only: Unrestricted			Voice only: Unrestricted			Voice plus text: Unrestricted		
Measure	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
Prior knowledge	.51	.16	52	.44	.18	55	.48	.17	53
Retention	.64[a]	.21	52	.61[a]	.20	55	.63[a]	.16	53
PT activity transfer	.88[a]	.32	52	.93[a]	.26	55	.89[a]	.32	53

Note. PT = Packet Tracer.
[a] Adjusted means using the pretest covariate.

Although the statistical modality effect with transfer scores in the first experiment was not replicated in the second experiment, which may have been due to a procedural error, the direction of the differences among the means was consistent in both experiments whereby the voice-only group scored higher than the text-only and voice-plus-text conditions. Thus, the results of Experiment 2, although not statistically confirming those of Experiment 1, suggest that perhaps a Type III error (where the direction of differences is incorrect) is unlikely.

ANOVAs were conducted to test for modality effects on the self-report items, with significant effects followed by Fisher's LSD post hoc tests and nonparametric Kruskal–Wallis H tests to control for Type I error inflation.

Similar to Experiment 1, statistical differences among the three conditions were found for three of the self-report cognitive load items. For the item, "Please rate your level of mental effort during the tutorial," $F(2, 157) = 4.91$, $MSE = 0.77$, $p < .05$, Cohen's $f = .25$ (medium), the voice-only tutorial ($M = 2.76$, $SD = 0.88$) was perceived as requiring less mental effort than both the text-only ($M = 3.27$, $SD = 0.84$) and voice-plus-text tutorials ($M = 3.15$, $SD = 0.91$). The Kruskal–Wallis H test confirmed the parametric results, $\chi^2(2, N = 160) = 9.72$. The item "Please rate the tutorial's level of complexity," $F(2, 157) = 3.21$, $MSE = 0.72$, $p < .05$, Cohen's $f = .20$ (medium), revealed a similar pattern. Students who had received the voice-only tutorial ($M = 3.27$, $SD = 0.89$) perceived the tutorial as less complex than both the text-only ($M = 3.67$, $SD = 0.76$) and voice-plus-text tutorials ($M = 3.57$, $SD = 0.88$). This finding was again confirmed by the nonparametric test, $\chi^2(2, N = 160) = 7.35$. For the item "Packet Tracer was frustrating," $F(2, 157) = 6.56$, $MSE = 1.19$, $p < .05$, Cohen's $f = .30$ (medium), students who had received the voice-only tutorial ($M = 2.78$, $SD = 1.01$) found PT to be less frustrating than students who received the text-only ($M = 3.29$, $SD = 1.04$) or voice-plus-text tutorial ($M = 3.55$, $SD = 1.22$). Again, the nonparametric test confirmed these findings, $\chi^2(2, N = 160) = 12.97$. Thus, it appears that the tutorial's text component negatively affected how students perceived PT. Taken together, the self-report item results from both experiments suggest that the voice condition caused students to experience less frustration and less extraneous cognitive load than the text condition.

General Discussion

The purpose for conducting this study was twofold. First, we sought to evaluate the combination of multimedia elements that is optimal for retention of tutorial knowledge and transfer of computer-networking skills. Second, we sought to empirically test the current theoretical models of multimedia instructional delivery by examining whether tutorial modality and restriction of a complex simulation's interface affected performance because of differences in demands on a learner's cognitive resources.

With regard to exploring how instructional modality delivery of the *My First PT Lab* tutorial affects performance on retention and transfer, we hypothesized that the voice-only condition would outperform both the text-only and voice-plus-text conditions because of the modality principle. Within the unrestricted conditions, this hypothesis was supported for the transfer test but not the retention test (statistically in Experiment 1). When PT's interface was unrestricted, the voice-only tutorial was better than the text-only one. This finding is consistent with both the modality effect and the redundancy principle (Mayer, 2001, 2005). In particular, Mayer (2001, 2005) has found that when presenting instructional content through multimedia animation, it is narration, rather than text or narration plus text, that leads to increased scores on measures of transfer but not retention.

The cognitive theory of multimedia learning explains this phenomenon. Voice with animation permits more efficient use of a learner's sensory input channels than text with animation. In using voice with animation, learners can more effectively attend to, select, and integrate relevant information from the content they hear and the content they view. Further, according to cognitive load theory, voice with animation may reduce extraneous processing in contrast to text with animation, which increases extraneous cognitive load because it splits attention between the two visual sources of information (Chandler & Sweller, 1992; Sweller, 1988). Thus, our finding that students in the voice condition performed better on transfer than those in the text condition is consistent with previous research on the modality effect.

We also explored the effect of restricting PT's interface. It was hypothesized that restricting the user's access to extraneous tools and objects on the interface would reduce distractions and lower the chance of getting lost within the simulation. Thus, the potential for extraneous cognitive overload would be reduced and learners would have additional cognitive resources available for germane processing. However, no differences were found between the unrestricted and restricted conditions on either retention or transfer measures. After observing the differences between the two conditions, this hypothesis may have been overly simplistic and perhaps other factors affected student reactions to the interface restrictions.

Certain features of the PT interface restriction manipulation may have actually negated any decrease in students' extraneous cognitive load because of two primary issues. First, the restricted condition had a third of the interface restricted; however, there were still numerous opportunities to get lost within the countless windows and options available in PT. Thus, the restricted condition was not so restricted that all extraneous options were removed. Second, students in the restricted group received numerous messages stating that the function they clicked on was locked. This message was presented with a loud bell noise. Perhaps the restriction messages discouraged or frustrated the learner (in the same way that receiving constant error messages increases frustration). According to the self-report data, students in restricted conditions did not report being more frustrated than students in the unrestricted conditions. Thus, perhaps some features may have lessened frustration, whereas others may have increased it. If the restriction messages had instead provided helpful content, such as specific instructions, instead of the *locked* message, then perhaps this feature may have been more helpful.

Finally, we also explored the possibility of an interaction effect between modality and PT's interface restriction. On the basis of existing literature (Mayer, 2001; Sweller, 1988), we hypothesized that the high-load conditions (text-only and voice-plus-text modalities and unrestricted conditions) would overload the learner's visual channel, thus resulting in poorer performance compared with the low-cognitive-load conditions (voice-only restricted), resulting in an additive effect of both voice-only modality and restriction and no interaction effect. Contrary to predictions, a Tutorial Modality × Interface Restriction interaction was found on the PT transfer measure in Experiment 1. Students in the unrestricted condition who received voice-only narration scored higher on the PT activity than students who received a text-based narration, whereas there was no modality effect within the restricted condition.

From a practical perspective, this finding within the unrestricted condition can be explained by the modality effect (Mayer, 2005). Students had to watch an animation of how to use PT while listening to or reading the narration. In the voice condition, they were able to listen to narration and watch the animation. In contrast, the text condition split students' attention in the visual channel between the text and the animation. Split attention effects have been documented and are part of the foundation of both the cognitive theory of multimedia learning and cognitive load theory (Paivio, 1986; Sweller, 1988).

It is less clear why this modality effect emerged in the unrestricted condition but not in the restricted condition. As mentioned previously, we suspect that rather than reducing extraneous cognitive load, some features of the restricted conditions may have actually increased extraneous load because students received *locked* messages with loud bell noises in their headphones. An *OK* button had to be clicked to close a *locked* message. Thus, the way the *locked* messages were delivered may have produced an unexpected annoyance that wiped out any benefits and prevented modality effects from emerging. On the other hand, the unrestricted conditions contained no *locked* messages. Thus, the latter group of students, although at risk for getting lost in extraneous activities in the simulation, also may have experienced less frustration with the loud bell noises. Another tenable explanation is that students who received optimal instruction (voice-only tutorial) were perhaps not led to click task-irrelevant buttons, whereas the text-only group may have gotten off task more often because of less-than-optimal instruction that then led to poor transfer performance. For this study, using parts of PT not needed for fixing the broken network would be considered off-task behavior. The voice-only tutorial may have allowed students to concentrate on just watching the animation of how to fix a network, whereas the text-only tutorial required students to read and watch, which may have limited their ability to focus on the animation.

Limitations

When we conducted this study, PT was not able to track the user's actions by creating log data files; thus, off-task behavior and problem-solving strategies were not measured. Such log files would have made students' cognitive processes observable and would have offered another element of variability between treatment groups. This study was also limited by its context, a lab setting, and by its participants, students with low prior knowledge of the subject matter. Such studies typically have a trade-off between the interacting contextual factors that impact real-world, pragmatic applications of the research and experimental control (Rieber, 2005). For instance, a challenge to the current study and to multimedia researchers in general is that the various measures were somewhat intrusive in delivery and may have contributed to an unnatural learning environment (DeLeeuw & Mayer, 2008). However, the experimental control in a lab setting such as this provided an opportunity to test the boundaries of the cognitive theory of multimedia learning and cognitive load theory. A logical sample for a future study would be actual Cisco NetAcad students. Further, any extension would benefit from increasing the relatively small sample size.

Conclusion

This study informs learning theory and instructional design by identifying optimal conditions for using complex simulations. Understanding how both modality and simulation interface restriction affect cognitive load can lead to the design of more effective educational software and improved learning. The results from this study have practical and theoretical implications.

Practical implications. Because this study used materials developed and currently used by the Cisco NetAcad, the findings may suggest instructional features that Cisco may explore to determine optimal methods for instructional delivery. Cisco currently uses the text-only tutorial and the unrestricted version of PT. Adding an audio narration to the tutorials has been considered by management but never implemented because of the cost of producing voice-over narrations in literally hundreds of languages. Yet, evidence from this study suggests that students may learn how to use complex simulations more effectively and efficiently and may perceive them to be easier to use if voice rather than text is used during the tutorial. Efficiency, from a cognitive load perspective, demonstrates students' decreased effort to achieve the greatest performance resultant of using materials that are optimally designed (Hoffman & Schraw, 2010). In terms of simulation interface restriction, there is no evidence that restricting is better or worse than not restricting. As was discussed earlier, it is possible that if these restricted messages were more constructive in terms of feedback and instruction, they could better facilitate learning rather than being a potential source of frustration. Of course, prior to making any changes or investments, more research is needed to replicate and extend the findings of this study. Our effort to replicate our findings in Experiment 2 was descriptively but not statistically consistent.

Theoretical implications. This study extends both the redundancy principle and the modality effect by examining the interaction between a complex simulation's interface and the modality of media presented during a tutorial showing how to use the simulation. The finding that within the unrestricted condition, the voice tutorial led to higher scores on the transfer test than the text tutorial did, has numerous theoretical implications. First, the cognitive theory of multimedia learning and, more specifically, the modality effect and the redundancy principle were extended in this study to a new context—a tutorial that teaches novices how to use a complex simulation. This adds further evidence to the importance of balancing presentation of instruction across visual and verbal channels so that the user's cognitive resources are not overloaded.

Future Directions

In terms of Mayer's modality line of research, a next step for this study would be to test Mayer and Johnson's (2008) revision to the redundancy principle. This would entail testing the difference between using the current voice-plus-text condition, in which the full text of narration is presented, and using a similar condition in which the text includes only a few keywords that direct the user's attention to critical parts of the animation. According to Mayer and

Johnson, the short redundant phrases condition would outperform the full text narration condition. Yet, Mayer and Johnson used content that dealt with processes, such as how lightning is formed, whereas this study first conveyed basic computer-networking information and then showed students how to use PT. Essentially, continued research is needed to determine the optimal modalities to be used for different types of multimedia learning environments and for different levels of learners (novices vs. experts). Restricting access to PT's interface also has theoretical implications for cognitive load theory. It was expected that restricting access to extraneous parts of PT's interface would reduce extraneous cognitive load; however, the user's affective reaction to repeated pop-up messages with a bell that indicated that the function was locked had not been taken into consideration.

More research is needed to determine the most effective methods for scaffolding through interface restriction. One potential method could be to simply gray out the buttons and functions that are restricted rather than popping up a *locked* message with a bell. Numerous software companies already use this method of not providing access to certain parts of the interface depending on how the software is being used. Specifically, future studies should examine how different types of restriction messages affect learning. If restricting access to parts of a simulation's interface is intended to scaffold the learner, then what is the optimal amount of restriction for various levels of learners? Should all extraneous functionalities be removed? Should the user be provided with explanatory feedback? Further research on restricting access to an interface is needed to identify how restriction can be utilized to lower extraneous cognitive load, provide feedback, and improve scaffolding.

Additional future directions for this particular line of research include examining the relationship between extraneous cognitive load and various usability measures. Also, measures of interest, measures of persistence, and behavioral tests of motivation are needed to shed light on how students' affective states influence their performance. For example, a behavioral test could include having students use PT and then take a short break. During the break, students can check their e-mail, surf the Web, continue using PT if they like, and so forth. A measure of whether students continue to use PT would be an indicator of motivation.

Future research should investigate the potential of using existing usability methods and measures for evaluating extraneous cognitive load within instructional technology environments. The undergraduate students in this study were generally novices in the field of computer networking; however, this study could be replicated with actual NetAcad students. The content and measures would need to assess more advanced levels of knowledge, and replicating this study with Cisco students may provide insight into the expertise reversal effect (Kalyuga et al., 2003).

References

Baddeley, A. D. (1986). *Working memory*. New York, NY: Oxford University Press.

Behrens, J. T., Frezzo, D. C., Mislevy, R. J., Kroopnick, M., & Wise, D. (2007). Structural, functional and semiotic symmetries in simulation-based games and assessments. In E. L. Baker, J. Dickieson, W. Wulfeck, & H. F. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 59–80). New York, NY: Erlbaum.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62,* 233–246.

Chang, K., Sung, T., & Chen, S.-F. (2001). Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning, 17,* 21–33. doi:10.1046/j.1365-2729.2001.00156.x

Clark, R. C., & Mayer, R. E. (2003). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning.* San Francisco, CA: Jossey-Bass.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100,* 223–234. doi:10.1037/0022-0663.100.1.223

Frezzo, D. C., Behrens, J. T., Mislevy, R. J. West, P., & DiCerbo, K. E. (2009). *Psychometric and evidentiary approaches to simulation assessment in Packet Tracer software.* In *ICNS 2009 Fifth International Conference on Networking and Services* (pp. 555–560). Retrieved from http://doi.ieeecomputersociety.org/10.1109/ICNS.2009.89

Frezzo, D., & Stanley, K. (2005, April). *Knowledge representations driving the design of computerized performance assessments in a complex simulated environment.* Paper presented at the annual meting of the American Educational Research Association, Montreal, Quebec, Canada. Retrieved from http://www.education.umd.edu/EDMS/mislevy/CiscoAERA2005/AERA_KR_FrezzoStanley.pdf

Gibson, D., Aldrich, C., & Prensky, M. (Eds.). (2006). *Games and simulations in online learning.* Hershey, PA: Idea Group.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15,* 313–331.

Gredler, M. E. (1992). *Designing and evaluating games and simulations.* London, England: Kogan Page.

Gredler, M. E. (2003). Games and simulations and their relationships to learning. *Educational Technology Research and Development, 21,* 571–582.

Herrington, J., & Kervin, L. (2007). Authentic learning supported by technology: Ten suggestions and cases of integration in classrooms. *Educational Media International, 44,* 219–236. doi:10.1080/09523980701491666

Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist, 45,* 1–14. doi:10.1080/00461520903213618

Kalyuga, S. (2007). Expertise-reversal effect and its implications for learner-tailored instruction. *Educational Psychological Review, 19,* 509–539. doi:10.1007/s10648-007-9054-3

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38,* 23–31. doi:10.1207/S15326985EP3801_4

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, England: Cambridge University Press.

Mayer, R. (2001). *Multimedia learning.* New York, NY: Cambridge University Press.

Mayer, R. (Ed.). (2005). *The Cambridge handbook of multimedia learning*: Cambridge, England: Cambridge University Press.

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100,* 380–386. doi:10.1037/0022-0663.100.2.380

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1,* 3–62. doi:10.1207/S15366359MEA0101_02

Nihalani, P. K., Mayrath, M. C., & Robinson, D. (2010). *When guidance* *harms and collaboration helps in computer simulation environments: An expertise reversal effect.* Manuscript under review.

Pavio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Penney, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17,* 398–422.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12,* 61–86. doi:10.1016/S0959-4752(01)00016-0

Rieber, L. (2005). Multimedia learning in games, simulations, and microworlds. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 549–568). Cambridge, England: Cambridge University Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257–285. doi:10.1207/s15516709cog1202_4

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159–168). Cambridge, England: Cambridge University Press.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12,* 185–233. doi:10.1207/s1532690xci1203_1

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251–296. doi:10.1023/A:1022193728205

van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17,* 147–177.

Received October 23, 2009
Revision received September 27, 2010
Accepted October 20, 2010 ■

Correction to Mayrath et al. (2011)

In the article "Varying Tutorial Modality and Interface Restriction to Maximize Transfer in a Complex Simulation Environment," by Michael C. Mayrath, Priya K. Nihalani, and Daniel H. Robinson (*Journal of Educational Psychology*, Advance online publication, January 24, 2011. doi: 10.1037/a0022369), the name of the author Laura G. Torres was omitted. Her name should appear after Nihalani and before Robinson. The other authors regret this error. All versions of this article have been corrected.