RESEARCH ARTICLE

Effects of organizational pictures and modality as a feedback strategy on learner comprehension and satisfaction

Albert D. Ritzhaupt[1] · Raymond Pastore[2] · Jiahui Wang[3] · Robert O. Davis[4]

Abstract One of the strongest instructional interventions documented in educational literature is the use of feedback to influence learning outcomes. However, there is lack of empirical research specifically pertaining to the use of multimedia in the feedback message. The purpose of this research was to test whether organizational pictures and modality as a feedback strategy had an effect on learner comprehension and satisfaction. The research design was a 2 Multimedia (Picture Present vs. Picture Absent) × 2 Modality (Narration vs. On-screen Text) × Trial (Trial 1 vs. Trial 2) with Multimedia and Modality serving as between-subject conditions and Trial serving as a repeated measure. One-hundred fifteen university students participated in the study. Findings show statistically significant increases in comprehension scores from Trial 1 to Trial 2 assessments for all four treatment conditions. Learners in the Picture present conditions were statistically more satisfied with the learning experience. Additionally, we used eye-tracking to verify the extent to which the pictures were used in the feedback message. A discussion and recommendations for future research and feedback design are provided.

Keywords Multimedia · Feedback · Multiple representations · Learner comprehension · Learner satisfaction · Modality · Eye-tracking

✉ Albert D. Ritzhaupt
  aritzhaupt@gmail.com

[1] School of Teaching and Learning, College of Education, University of Florida, 2423 Norman Hall, P.O. Box 117048, Gainesville, FL 32611, USA

[2] University of North Carolina Wilmington, Wilmington, NC, USA

[3] University of Florida, Gainesville, FL, USA

[4] English Literature and Language Technology Department, Hankuk University of Foreign Studies, Seoul, South Korea

Introduction

The use of feedback in instruction is not a new idea. Instructional designers, teachers, and trainers routinely incorporate feedback into instruction as a learning strategy in order to improve learning and performance. For example, feedback is tied to many instructional strategies, such as Gagne's Nine Events of Instruction (Gagne 1985), which is the backbone of many courses and educational programs. Feedback has further been identified as one of the strongest instructional strategies in meta-analytic research on educational practices (Hattie 2008; Hattie and Timperley 2007). Current research and practice has shown that feedback as a strategy is an effective means to improve student learning and performance, particularly in computer-based learning environments (Kollöffel and de Jong 2015; Van der Kleij et al. 2015). However, there are not clear recommendations on how to best design feedback in order to improve learning and decrease load on working memory (Ritzhaupt and Kealy 2015; Shute 2008). Very few studies have examined how different types of feedback representations affect learning and performance in the context of a multimedia learning environment (Narciss 2008; Van der Kleij et al. 2015). As a result, the current study aims to examine the types of representations (verbal and nonverbal) in feedback during the learning process to make recommendations to better design feedback for computer-based and multimedia learning environments. This unexplored area of research can provide useful recommendations to instructional designers and teachers for harnessing the power of feedback.

Multimedia representations

When designing instruction, one must consider the types of representations, such as, verbal and/or non-verbal, to be presented to the learners (Ainsworth and Loizou 2003). The type of representations selected affects the learning process and the cognitive load placed on working memory. As a result, it is hypothesized that feedback carries these same limitations; however, very few experimental studies have examined if this holds true. The idea that the types of representations presented during instruction affect achievement and cognitive load stems from Dual Coding Theory (DCT). DCT emphasizes that our working memory is comprised of two memory channels (verbal and non-verbal), which each have a certain capacity and can process information separately and independently from one another (Paivio 1991). When connections between the verbal and non-verbal channels are made, cognitive load on working memory is decreased, which can improve learning outcomes (Paas and Ayres 2014).

Consequently, instructional designers strive to utilize multiple representations in instruction in order to better utilize the load on working memory in order to increase learner achievement. This idea has generated numerous theories aimed at creating better instruction, chiefly among them, the Cognitive Theory of Multimedia Learning (CTML). CTML is based on three assumptions (1) working memory is made up of a dual modality input channel system, (2) working memory capacity is limited, and (3) that learners engage in active processing (Mayer 2005). In light of CTML, Mayer and colleagues developed a number of design recommendations aimed at creating instruction which should theoretically decrease working memory load and improve learning outcomes. These recommendations from CTML stem from the idea that multiple representations are better for learning than just one, otherwise known as the multimedia principle. The multimedia principle states that learning from pictures and words leads to higher level learning outcomes than words alone. Research studies have provided evidence in support of this theory (Eitel et al.

2013; Mayer 2009). For instance, Eitel et al. (2013) conducted a study comparing text and image groups to just text groups on high- and low-level knowledge. University students were exposed to a computer-based instructional treatment. The multiple representation groups scored higher on the high-level knowledge test. However, not all representations are utilized the same in working memory. Some combinations may use more or less cognitive resources than another (Mayer 2014). As a result, a multitude of additional hypotheses were developed from the CTML to explore the conditions of the multimedia principle. For further reading on Mayer and colleagues principles, see Mayer (2005, 2014) as only the split attention and modality principles will be discussed in this paper.

When learners examine multiple representations consisting of words and pictures, learners need to hold the concepts from the words in working memory so that they can examine the pictures. This is referred to as representation holding and requires learners to split their attention between two representations (Sweller et al. 2011). Representation holding can increase learners' cognitive load, causing an increasing burden on their working memory (Mayer 2014; Mayer and Anderson 1992). Therefore, the modality principle was developed. The modality principle suggests that when using verbal narration instead of on-screen text, one can focus on the picture while listening to the audio (generally expressed as narration). Thus, representation holding does not occur when simultaneously presenting the picture and words via narration. Cognitive load is decreased and learning is increased when compared to on-screen text and picture representations. This finding has been confirmed in a number of studies (Scheiter et al. 2014; Pastore 2012; Fiorella et al. 2012). For instance, Herrlinger et al. (2016) placed participants into groups consisting of images with either spoken or written text. Results of the study revealed that participants presented with spoken, rather than written text, saw an increase in learning outcomes. Two experiments conducted by Mayer and Moreno (1998) using information on the process of lightning, and how a car's braking system works, showed that students learned better when verbal information was presented in narration as opposed to on-screen text. Thus, the type of representation can affect load and performance. Having said that, the modality effect has only been tested on initial instruction. The question remains as to whether this will apply to other parts of instruction (i.e., feedback).

Feedback in instruction

Feedback is a vital element to support the process of learning (Cohen 1985; Gagne et al. 1992; Hattie and Timperley 2007; Shute 2008). The value of feedback has been recognized in learning paradigms such as behaviorism, cognitivism, constructivism, and open learning environments (Mory 2004). This multidimensional view of feedback has caused it to be defined a variety of different ways. Gagne (1985) speaks of feedback providing confirmation, or a level of correctness, to a learner's performance. This definition implies a performance of some type that is correct or needs to be corrected. Hattie and Timperley (2007) have defined feedback as "information provided by an agent (e.g., teacher, peer, book, parent, self, experience, computer) regarding aspects of one's performance or understanding" (p. 81). This definition of feedback is broader than the previous as understanding becomes a component that needs to be altered and can happen through various agents. In computer-based learning environments, feedback is seen as any message provided to the learner after a task (Wager and Wager 1985). This can be in the form of statement delivered on-screen, animated graphic, or sound to alert the learner a response is either correct or incorrect. Because of the rapid advancement of technology, we are able to

test various types of feedback through different methods and combinations of multimedia (e.g., animated pedagogical agents).

As previously noted, the modality principle suggests that words presented as audio narration rather than text are better for learning when presented in a multimedia environment (Mayer 2014; Mayer and Moreno 2002). The origin of this principle comes from classic verbal learning research on short-term memory (Moreno 2006). Penney (1989), who proposed the separate streams hypothesis, stated, "in short-term memory tasks, auditory presentation almost always resulted in higher recall than did visual presentation" (p. 398). According to Penny, information is processed in either the Acoustic code or the Phonological code. When words enter short-term memory through the auditory modality, the information interacts with the Acoustic code and the Phonological code simultaneously. Without any interference, this information can be maintained for up to a minute (Engle and Roberts 1982). However, the visual modality of words in short-term memory can be disrupted by concurrent narration because the visual code and the Phonological code are being interfered upon by the audio.

Research has shown evidence that support the modality principle's empirical claims as described in the previous section (Atkinson 2002; Mayer and Anderson 1992; Moreno and Mayer 1999; Fiorella et al. 2012). However, for all the experiments conducted on the modality principle, there have been a lack of empirical studies on feedback within the modality principle (Ritzhaupt and Kealy 2015; Johnson and Priest 2014). For example, O'Neil et al. (2010) examined feedback after team collaboration tasks. One team was given text only feedback, while the other team received text and narration feedback. The team receiving feedback from text and narration had better results than the team receiving text only feedback. Another study by Fiorella et al. (2012) placed participants into groups consisting of either visuals with narrative (spoken words) feedback, visuals with printed feedback, or visuals with no feedback. Results of the study revealed that the visual with narrative (spoken words) feedback group scored significantly higher than the other groups on both low- and high-level comprehension tests.

Feedback with pictures

The lack of research examining the use of pictures as a feedback strategy is surprising. Modern technology (e.g., authoring systems or Learning Management Systems) supports the use of pictures in instruction and the cost of producing relevant imagery for instruction has dramatically declined in recent years (Morrison et al. 2010). Previous studies have shown that pictures can help provide assistance to access and understand words (Schallert 1980), and to improve students' learning of words when pictures are present (Carney and Levin 2002; Levin et al. 1987). The limited research that has been conducted on feedback with pictures has provided some insight on the value of pictures in this form. Moreno and Valdez (2005) conducted research that showed words and pictures together scored significantly higher in retention and transfer of information than words or pictures alone. Ritzhaupt and Kealy (2015) conducted a study to examine the use of multiple representations in feedback. Although their research was unable to pinpoint a condition in which pictorial feedback was superior to text-alone feedback, their research did have some shortcomings and recommendations. Chiefly, their research recommended the use of different types of pictures in the initial instruction and the feedback cycle. Levin (1981) suggests pictures can serve as decorational, representational, organizational, and transformational. While decorational images serve no purpose and can actually hinder the learning process, organizational pictures have been found to have moderate effects on

learning (Carney and Levin 2002; Levin 1981). Ritzhaupt and Kealy (2015) explicitly used representational pictures in their interventions. Thus, one of their primary recommendations was to employ feedback treatments that used more sophisticated pictures, such as organizational pictures that typically manifest themselves as maps or diagrams. Conceivably, using organizational pictures might provide the additional support to learners necessary to have positive outcomes on learning and performance. Further, Ritzhaupt and Kealy (2015) also did not make use of the modality principle in their interventions as all used on-screen text in the treatments. Their recommendations were to implement a study that operationalized the modality principle with different types of pictures.

Eye-tracking research

The eye-mind hypothesis (Just and Carpenter 1980) suggests that eye movement recordings can provide a trace of where the person's attention is directed to and where the person is cognitively processing as well. Therefore, compared to traditional outcome measures such as tests of learning comprehension and transfer, eye-tracking method could provide valuable information on the underlying visual and cognitive processing that occur during learning. Historically, the eye-tracking technique has been widely used to study reading (Rayner 1998; Schneps et al. 2013) and scene viewing (Henderson et al. 1999). Visual attention is typically measured in the form of fixations, which occurs when an eye settles on something for around 300 ms. Fixation duration is longer on a difficult or unfamiliar word during reading (Rayner 1998) and an aspect of the problem that requires more cognitive processing will receive more and longer fixations (Carpenter and Shah 1998). In recent years, eye-tracking method started being employed to study multimedia learning (Mayer 2010) and eye-tracking has offered various possibilities for research on multimedia learning and instruction (van Gog and Scheiter 2010; Wang and Antonenko 2017). Specifically, eye movement data has shed light on the well-known multimedia learning principles such as the spatial contiguity principle (Johnson and Mayer 2010; Holsanova et al. 2009) and signaling principle (Ozcelik et al. 2010). Moreover, by tracking individual learners' eye movements, more could be understood as to how learners with different levels of prior knowledge or expertise perceive and interpret dynamic visualizations (Jarodzka et al. 2010) and offer insights on how multimedia learning materials can be adapted to accommodate individual needs.

To this point, very few studies have used the eye-tracking technologies to investigate how learners process pictorial and textual visual stimuli in multimedia learning context, especially during the feedback cycle of learning (Hegarty and Just 1993; Schmidt-Weigand et al. 2010). Eye-tracking studies in this regard contribute to our understanding of how people process visual and textual information, which have found learners generally adopted a text-directed processing strategy (e.g., Hegarty and Just 1993; Schmidt-Weigand et al. 2010). In a pioneering study, Hegarty and Just (1993) investigated learning from diagrams and verbal descriptions of pulley systems. Participants' eye movement patterns indicated that they read the text about a specific component of pulley system before inspecting its referent in the diagram. Similarly, Schmidt-Weigand et al. (2010) examined visual attention distribution in learning from text and pictures under system-paced and self-paced conditions. Their study found while presented with multimedia instruction on the formation of light, learners who are presented with on-screen texts and visualization spent more time reading the texts than inspecting the visualization. Moreover, learners who listened to spoken texts attended to visualizations more, compared to those who read the on-screen texts. Also, learners who listened to spoken texts performed significantly better

in the follow-up visual memory test (e.g., sketch a picture of how electric charges arise in a thundercloud). These findings suggested learners' processed multimedia stimuli in a text-oriented manner and spoken texts leave more cognitive resources available for attending to visualizations, thus leading to an enhanced memory of visualizations. No study, to our knowledge, has adopted the eye-tracking technique to study how learners process words and pictures used in the feedback message. The question of interest to the present study is whether learners also adopt a text-oriented strategy to process picture feedback and if narrated text in feedback can allow more cognitive resources for processing visualizations.

In addition to fixations on text and picture respectively, researchers have also examined integrative transitions between text and picture to gauge learner's efforts in integrating textual and pictorial information (Mason et al. 2013; Johnson and Mayer 2012; Holsanova et al. 2009). Integrative transitions occur when learners' eye fixation moves from text to picture or from picture to text. For example, Holsanova et al. (2009) found that readers made more integrative transitions in the integrated format (8.2%) than the separated format (2.3%), possibly because integrative format with a shorter distance between text and picture facilitates integration. Further, as noted in Ritzhaupt and Kealy (2015), the researchers were unable to conclusively document the extent to which the learners attended to the pictorial information. For the current study, the eye-tracking technique could inform the attentional dynamics that occur as learners integrate picture and text during the feedback cycle of learning.

Purpose

Thus, the purpose of this research was to extend the research by Ritzhaupt and Kealy (2015) to address the recommendations and shortcomings in their research. Specifically, the current research uses organizational pictures of the human heart and a complex scientific explanation as the initial instruction and feedback cycle intervention. Equally important, the present research operationalizes the modality principle in the treatment conditions, making use of both channels (auditory and visual) to best influence learning outcomes, and satisfaction. Further, this study uses eye-tracking method to examine the extent to which the pictorial information is used by learners when they fixate on the feedback message. The overall purpose of this research is to identify conditions in which the use of multimedia and modality are appropriate for the design of effective and efficient feedback in computer-based learning environments. Our primary research question is what influence do the multimedia and modality principles have on learner comprehension and satisfaction during the feedback cycle of learning? Our secondary research question is to what extent are the learners using the pictures during the feedback cycle?

Method

Design and participants

The experiment was a 2 Multimedia (Picture Present vs. Picture Absent) × 2 Modality (Narration vs. On-screen Text) × Trial (Trial 1 vs. Trial) with Multimedia and Modality serving as between-subject conditions and Trial (feedback cycle) serving as a repeated measure. Participants ($N = 115$) were recruited from two public, southeastern universities

in the United States after making prior arrangements with course instructors. Participants were randomly assigned to a treatment group resulting in the following distribution: PN = 31, PO = 27, AN = 25 and AO = 32. This distribution is shown in Table 1.

Sixty-four percent of the participants were female. Seventy-eight percent of the participants classified as White. Ten percent of the participants were freshmen, 23% sophomore, 38% juniors, 10% seniors, and the remaining participants were either graduate students or other. The ages of the participants ranged from 18 to 44 with an average age of 23.31 (SD 4.73). All participants were enrolled in educational technology courses offered by a College of Education. However, participants were enrolled in a wide variety of majors. An important note is that only a subset of 20 participants were invited to complete the eye-tracking component of the current research study.

Instructional materials

The 2000-word script was originally developed by Dwyer (1965) and later revised by Dwyer and Lamberski (1983). It focused on the physiology and function of the human heart and included 19 static line drawings with color shaded regions. This content was designed by an instructional designer, reviewed by content experts, and piloted before initial use (Dwyer 1965, 1978; Dwyer and Lamberski 1983). The Kincaid–Flesh reading grade level was 9.3.

Eye-tracking apparatus

For the subset of 20 participants that completed the eye-tracking component of the study, the human heart tutorial and follow-up assessments were displayed on an external 20-inch flat panel monitor viewed at a 55-cm distance, with a resolution of 1600 by 900 pixels and a refresh rate of 60 Hz. Participants used a chinrest (SR-HDR) with a forehead bar to minimize head movement. Eye-tracking data was collected via Eyelink 1000 Plus system (SR Research, Ontario, Canada) using a desktop-mount (see Fig. 1). Participants used a Bluetooth mouse to proceed with the tutorial and responded to the follow-up assessment in the form of multiple-choice questions. Screen Recorder software was used to simultaneously capture locus of participants' gaze while recording the screen activity, at a sampling rate of 1000 Hz.

Criterion measures

Learner comprehension

The comprehension test was designed to measure students' transfer of problem solving and consisted of 20 multiple-choice questions. "Thus in order to perform well on this portion of the test, students had to have an understanding of facts, concepts, rules/procedures, and

Table 1 Distribution of participants by treatment condition		Modality	
		Narration	On-screen Text
Picture	No picture	25	32
	Picture	31	27

Fig. 1 Experimental setup for eye tracking component of picture feedback study
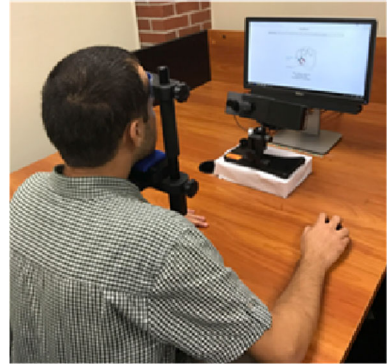
problem solving objectives. This test required students to thoroughly understand the heart, its functions, and processes in both the systolic and diastolic phases" (Pastore 2010). This measure was originally developed by Dwyer (1978) and had been analyzed by Dwyer (1978) in over 100 experimental studies and produced a Kuder–Richardson Formula 20 (KR-20) reliability score of 0.70. The KR-20 for Trial 1 was 0.66 and for Trial 2 was 0.76. Two example learner comprehension items are shown in Table 2.

Learner satisfaction

The learner satisfaction survey consisted of 9 items from previous studies of multimedia learning environments (Ritzhaupt and Barron 2008; Ritzhaupt et al. 2011). The instrument uses a five-point scale with two bipolar adjectives on both sides. For instance, on the left-most side is the word ''obscure'' and on the right-most side is the word ''clear.'' The items were designed to measure a learner's satisfaction with the intervention. The Cronbach's α for the satisfaction survey was high at $\alpha = 0.91$.

Computer programs

The four conditions and criterion measures were created using Captivate 8.0, PHP, MySQL, CSS, and HTML. PHP and MySQL were used to capture the learner responses to the criterion measures while CSS and HTML were used for the presentation of this information in correspondence with the Captivate 8.0 tutorials. Captivate 8.0 was used to build the interventions based on the instructional materials presented. A sample screen shot

Table 2 Learner comprehension test sample questions

| When blood is being forced out the right ventricle, in which position is the tricuspid valve?	When the blood in the aorta is exerting a superior pressure on the aortic valve, what is the position of the mitral valve?
A. Beginning to open	A. Closed
B. Beginning to close	B. Open
C. Open	C. Beginning to open
D. Closed	D. Confined by pressure from the right ventricle

from the Picture and On-screen text (PO) condition is shown in Fig. 2, which illustrates the look-and-feel of the initial instruction. Figure 3 shows the feedback cycle from the same condition for one of the comprehension assessment items from Trial 1. Please note that in the narration conditions, the same text was spoken by an English speaking male's voice.

An important note about our intervention for the feedback cycle was that we did not include response correctness (e.g., correct or incorrect) as a variable in the instructional message. Although response correctness is an important variable in computer-based learning environments, it also serves as an unnecessary variable in the intervention. Prior research has documented that learners use feedback without response correctness to self-regulate their learning (Kealy and Ritzhaupt 2010). Further, response correctness alone has been shown to have weak effects on learning outcome measures (Van der Kleij et al. 2015). Thus, we opted to not include response correctness as a variable in our feedback intervention, which would have duplicated the number of conditions in the study by two. Further, the verbal feedback messages were intentionally designed to be short. Longer elaborations can cause learners to lose focus of the relevant content (Shute 2008).

Procedures

Prior to beginning the *Human Heart Tutorial*, each participant completed the background survey. After, the *Human Heart Tutorial* was randomly assigned based on one of the four treatment conditions for the initial instruction and feedback intervention. Each participant was automatically assigned a unique sequential integer, and the computer programs used modulus arithmetic to randomly assign each participant to a different Picture treatment (Present vs. Absent) and Modality condition (On-screen text vs. Narration). Participants were not informed of which condition they were assigned, and the researchers were also unaware (a double-blind random assignment). Trial 1 was administered by item with the
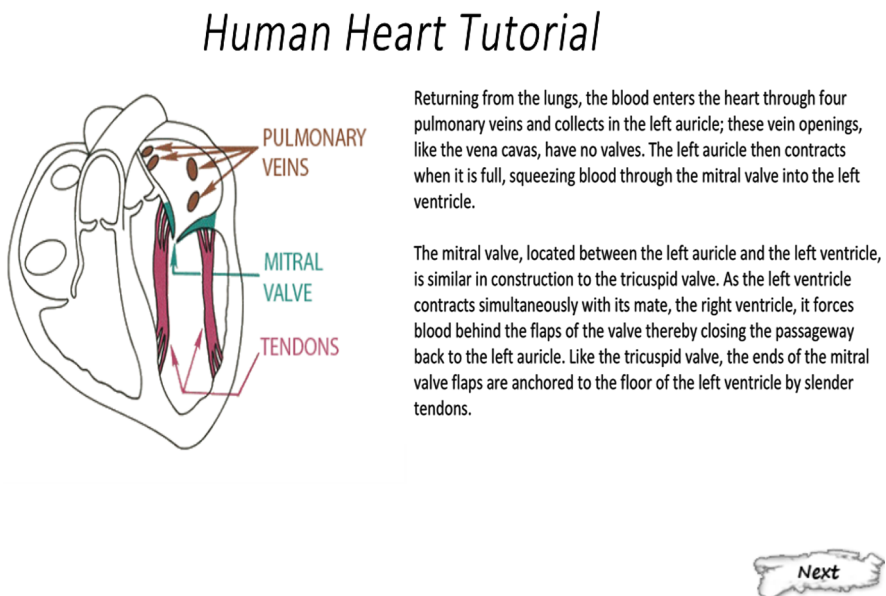
Fig. 2 Sample screenshot of initial instruction in the Picture and On-screen text (PO) condition

Fig. 3 Screenshot of feedback
from the Picture and On-screen
text condition (PO)

TRICUSPID
VALVE

TENDONS

The mitral valve is
most like the
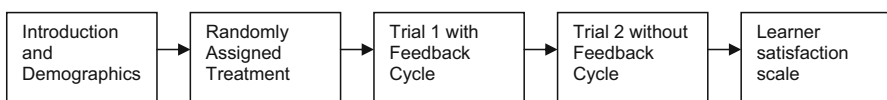tricuspid in function
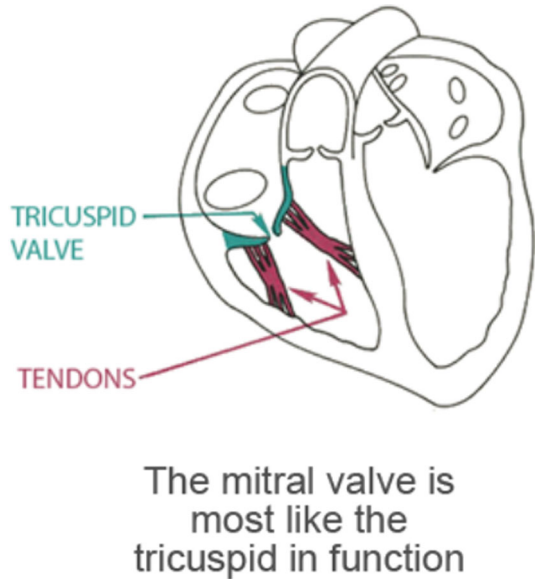
associated feedback treatment in one of the four conditions, and intended to demonstrate
the fidelity of the feedback with the expectation that comprehension would improve on the
Trial 2. After completing the *Human Heart Tutorial* and Trial 1 assessment with feedback,
participants responded to a Trial 2 performance assessment and the satisfaction survey. The
satisfaction survey was designed to measure learners' attitudes toward the intervention.
Figure 4 illustrates the sequence of the research study. The average time to complete the
full intervention was approximately 30 min.

Eye-tracking procedures

The same procedures were followed as noted above with a few key differences noted in
this section. At the beginning of the experiment, the gaze of each participant was calibrated
and validated with a 13-point calibration algorithm. Calibration was repeated until accu-
racy criteria "good" was satisfied. While participants watched the tutorial and responded to
the assessment items, their eye movements were simultaneously recorded. After watching
the tutorial on the human heart, participants responded to 20 multiple-choice questions,
each followed by item-level feedback in one of the four conditions (i.e., PO, PN, AN, AO).

Data analysis

Prior to inferential analysis, comprehension and satisfaction scores were summated to form
composites for the learner comprehension and satisfaction measures, which served as the

| Introduction and Demographics | → | Randomly Assigned Treatment | → | Trial 1 with Feedback Cycle | → | Trial 2 without Feedback Cycle | → | Learner satisfaction scale |

Fig. 4 Research intervention sequence

dependent measures in this study. There are three assumptions of an Analysis of Variance (ANOVA), which include normality, homogeneity of the variance, and independence of observation (Stevens 1990). A Levene's test was used to test for the assumption of homogeneity of the variance, and the skewness and kurtosis were used to evaluate the normality assumption. The data are assumed to be independent because of the methodical assignment procedures as described in the method.

Results

There were no severe departures from normality for the learner comprehension measure as evidenced by the skewness and kurtosis for Trial 1 and Trial 2 performance were 1.32 and 2.27, and 0.83 and 0.19, respectively. The Levene's test was executed on the Trial 1 and Trial 2 assessments at $F(3,111) = 1.10$, $p = 0.35$ and $F(3,111) = 0.44$, $p = 0.73$, which provides evidence that the error variance of the measures are equal across the conditions. Thus, our data appeared to be well-suited for ANOVA.

Learner comprehension

The learner comprehension data were entered into a 2 Picture × 2 Modality × 2 Trial repeated measures ANOVA with both Picture and Modality serving as between subjects conditions and Trial serving as a within subjects condition. The results indicate that Trial was statistically significant $F(1,111) = 49.32$, $p < 0.01$, partial $\eta^2 = 0.31$. As anticipated, learner comprehension after the four different feedback treatments improved significantly from Trial 1 to Trial 2 with approximately 31% of the variability explained in the model. The mean scores and standard deviations for learner comprehension on Trial 1 and Trial 2 are shown in Table 3 by the four treatment conditions. As can be gleaned, learner comprehension increased for all four treatment conditions from Trial 1 to Trial 2 on the assessments.

When investigating the effects of Picture conditions (Present vs. Absent) in isolation, the results indicated the absence of a statistically significant main effect for Picture $F(1,111) = 0.29$, $p = 0.59$, partial $\eta^2 = 0.03$. However, the means and standard deviations between groups on Trial 2 performance did show mild variation (see Table 3). The results also indicated there was no significant interaction effects between Trial x Picture at $F(1,111) = 3.66$, $p = 0.06$, partial $\eta^2 = 0.03$. This indicates that Picture did not interact with Trial. However, it is notable that this interaction effect is approaching statistical significance ($p = 0.06$). That is, there may be an effect here worth examining further in larger sample size replication studies that command higher statistical power and overall generalizability. This interaction effect is key to the hypothesis that multimedia feedback is more effective than text-alone feedback.

Table 3 Mean and standard deviations of learner comprehension scores by treatment condition and trial

	Trial 1		Trial 2	
	Narration	On-screen Text	Narration	On-screen Text
Picture Present	6.94 (3.37)	7.44 (4.34)	8.19 (4.37)	9.59 (3.85)
Picture Absent	6.36 (2.94)	6.06 (3.02)	9.60 (4.48)	8.78 (4.01)

We also did not detect a significant main effect for the Modality (Narration vs. On-screen text) across the four treatments at $F(1,111) = 0.10$, $p = 0.76$, partial $\eta^2 = 0.001$. It appears that Modality of the learning materials did not materialize in the interaction effect with the Trial either at $F(1,111) = 0.80$, $p = 0.78$, partial $\eta^2 = 0.001$. In examining the remaining interaction effects, we did not observe a statistically significant interaction between Picture and Modality at $F(1,111) = 1.41$, $p = 0.24$, partial $\eta^2 = 0.01$. Nor did we detect a significate three-way interaction among Picture, Modality, and Trial at $F(1,111) = 1.12$, $p = 0.29$, partial $\eta^2 = 0.01$.

Learner satisfaction

The learner satisfaction scores were entered into a 2 Picture (Present vs. Absent) × 2 Modality (Narration vs. On-screen text) ANOVA with both variables serving as between-subject conditions. We did not detect a statistically significant main effect for Modality at $F(1,111) = 0.04$, $p = 0.84$, partial $\eta^2 = 0.00$. However, we did detect a statistically significant main effect for Picture at $F(1,111) = 14.24$, $p < 0.01$, partial $\eta^2 = 0.11$. Further, the Picture conditions explains approximately 11% of the variability in the model. That is, the learners in the Picture present conditions were statistically significantly more satisfied with their experience than their Picture absent counterparts. The means and standard deviations for the four treatment conditions are shown in Table 4. As can be gleaned, the scores for the Picture present conditions are higher for both Modality conditions.

Eye-tracking results

In order to explore how participants attended to the word and picture components while provided with multimedia feedback message, we examined the eye-tracking data generated from participants who have been assigned to PN (Picture Present and Narration) and PO (Picture Present and On-screen Text) conditions. For both PN and PO conditions, we defined the areas of picture on the 20-item Trial 1 assessment, and the associated feedback pages as picture Interest Areas (IA) across all participants, with the size of interest areas slightly varied across feedback messages. Also, we defined the areas of text as text IA for PO condition. The descriptive statistics on total fixation time and number of fixations are summarized in Table 5. Results indicated that participants allocated more fixations to the picture in the PN condition ($U = 115.00$, $z = 8.846$, $p < = 0.001$) and longer total fixation time to the picture in the PN condition as well ($U = 220.00$, $z = -8.293$, $p < 0.001$). This finding suggested learner's viewing behavior on the pictorial feedback is largely influenced by text modality (Narration vs. On-screen text).

Also, results indicated that in the PO condition, participants spent significantly longer time fixating on the text compared to the picture ($U = 736.00$, $z = -5.585$, $p < 0.001$),

Table 4 Mean and standard deviations of learner satisfaction scores by treatment condition

	Narration	On-screen Text
Picture Present	22.97 (8.73)	22.26 (7.77)
Picture Absent	16.64 (4.06)	17.94 (8.11)

Table 5 Mean and standard deviation of number of fixations on IA and total fixation time on IA

Conditions	PO		PN	
	M	SD	M	SD
Number of fixations on IA				
Picture*	3.85	6.60	17.45	6.15
Text	6.98	6.01	–	–
Total fixation time on IA (ms)				
Picture*	1021.52	1951.02	4662.42	1982.60
Text	1703.57	1487.50	–	–

* Indicates significant difference at $p < 0.05$

and a larger number of fixations on the text ($U = 443.00$, z $= -7.125$, $p < 0.001$). Furthermore, the number of transitions between picture and text IAs was examined for PO condition to explore the split attention effect between text and picture when they both were present. The results indicated participants did look back and forth between text and picture IAs and they performed more transitions from picture to text (M 2.86, SD 4.74) than from text to picture (M 1.46, SD 4.56), $U = 5379.00$, z $= -3.665$, $p < 0.001$).

Discussion

Interpretation of the results must be viewed within the limitations and delimitations of the present study. It was assumed that research participants did not have hearing impairments that might render the audio treatment interventions unintelligible and did not have extensive previous experience with the human heart functions. The results of this study should not be generalized outside of the population of university students in higher education or populations with similar demographics. The type of content employed in this study would likely be characterized as high intrinsic cognitive load (Chandler and Sweller 1991) as opposed to educational materials that pose less intrinsic cognitive load on the learners' working memory during instruction and feedback. Finally, we did not collect any qualitative data in our data collection procedures, and consequently, cannot triangulate our data sources. In light of these limitations and delimitations, the study resulted in several important findings.

Our primary research question was answered by the analysis of learner comprehension and satisfaction. Analogous to the findings of Ritzhaupt and Kealy (2015), we were unable to identify a statistically significant condition on the learner comprehension measure in which the two independent variables in the present study—Picture and Modality—were superior on the learner comprehension dependent measure. However, one key finding is that the interaction effect between Trial and Picture was approaching statistical significance ($p = 0.06$). This is an important finding in that the interaction with the feedback cycle (Trial) and Picture condition (Present vs. Absent) would suggest the multimedia principle was present, but not strong and durable enough in the treatment conditions. This might be explained by two related concepts in multimedia learning.

The intervention used in the present study, although rigorous and relevant, made use of a complex formal language in the delivery of the verbal materials. Specifically, we violated two of the common multimedia principles by using these previously studied instructional materials by Dwyer (1965): coherence and personalization. The coherence principle suggest that extraneous words and pictures should be excluded from a multimedia presentation

(Mayer 2009) rather than included. Although the Kincaid–Flesch was at a 9.3 grade level, there were seductive details included in the verbal presentation that may have overloaded working memory or confused the learners. Seductive details are unnecessary pieces of information about the subject-matter that can overload working memory. Second, the personalization principles suggests that learners attend to words better in a conversational style as oppose to a formal style (Mayer 2009). The *Human Heart Tutorial* undoubtedly used a formal style of language, which again, may have overloaded working memory and deteriorated from the learning outcomes.

However, the learning comprehension measure is only one side of the tricky equation of learning outcomes and studying the efficacy of multimedia interventions. We did detect a statistically significant main effect in favor of the Picture present condition on learning satisfaction. Precisely, learners in the Picture present condition were statistically more satisfied with the learning experience than those without the pictures. Thus, although we do not have conclusive evidence of the role of the multimedia principle in feedback, the evidence that we do have suggests that relevant organizational pictures should be used appropriately as a feedback strategy in quality computer-based learning environments.

The results do not show any evidence in support of the modality principle in feedback. Again, this finding might be attributable to the lack of using the Coherence and Personalization principles in the interventions. Additionally, even though statistical significance was not found, the means of the picture present conditions are not in-line with the modality principle; rather, they point to the reverse. As a result, this finding should be further examined in future research so that the modality principle in a feedback environment can be better understood. A study by Cheon et al. (2011) revealed a reverse modality effect wherein those studying visual text outperformed those studying spoken text in their intervention. We may not have fully documented all conditions in which the modality principle works best.

Our second research question is addressed through the eye-tracking component of the study. This is perhaps the first study to incorporate eye-tracking methodology into the analysis of the feedback message in instruction. One consistent problem by Ritzhaupt and Kealy (2015) is that we did not know the extent to which the learners are attending to the pictorial content in the interventions. Eye-tracking technique adopted in the current study illustrated the attentional dynamics as learners' processed multimedia feedback. One important finding is learners allocated significantly more attention to the texts when pictorial and textual information were simultaneously present. This observation provides more support for a text-oriented strategy in processing multimedia instruction (e.g., Hegarty and Just 1993). It is noteworthy that learners devoted more fixations to the picture when text was narrated rather than displayed on the screen, which converges with previous findings from Schmidt-Weigand et al. (2010). However, our study did not replicate the modality effect—those who listened to the narrated text did not outperform in the comprehension test. A possible explanation is that although those learners had more opportunities for examining the picture in the feedback cycle, this extra effort did not exert enough influence on the comprehension task. It is reasonable to speculate that the textual information, either displayed on screen or narrated, is sufficient for understanding how human heart works, which is the essence of the comprehension task.

With the increased popularity of multimedia being incorporated into educational programs, coupled with the ease of creating and using pictorial technology (e.g., Photoshop), underscores the importance of this research and future research in this area. Learners are using content that is often absent of relevant and meaningful multimedia feedback to enhance the learning outcomes. We must fully understand the influence of this form of

instruction on learning. At minimum, this paper contributes to this understanding and opens further areas for exploration.

Recommendations for future research

As our research leaves the conversation about the use of multimedia in feedback unresolved, we provide some recommendations for future researchers in this area. First, we need a whole branch of studies on the effectiveness of multimedia feedback in terms of replication studies to lead to effective design principles in this research area. Our findings and those of Ritzhaupt and Kealy (2015) are a starting point for this line of research. Future research, after a solid number of studies on multimedia feedback have been examined, might lead to a meta-analytic study to documents and generalize to the larger population of learners and potential moderators.

Second, future multimedia feedback studies should aim to incorporate more of the design principles derived from the Cognitive Theory of Multimedia Learning (CTML) by Richard Mayer and his colleagues. As already noted, our instructional materials did not operationalize the principles of Coherence and Personalization. This design flaw may be an explanation for the modest multimedia principle finding, and the disappointing finding related to the modality principle.

Recommendation on designing feedback

Our present study indicates a few notions related to the design of effective multimedia feedback messages. We have demonstrated that the verbal message should be short and on target to the information the learner needs to improve their learning and performance. It is unnecessary for elaborations to dwell on seductive details in order to influence learning. The pictorial information (e.g., pictures, animation, etc.) should complement and be semantically related to the verbal message. Again, although the multimedia principle was not statistically significant in this research, the p value was approaching significance. Further, the learner satisfaction measures suggest that learners preferred instruction with relevant pictures. Also, verbal feedback can be provided as narration in future designs, as it could allow learners to devote more visual attention to pictorial feedback without sacrificing the comprehension of verbal information.

The use of immediate feedback in computer-based learning environments is a useful design strategy for both high and low intrinsically cognitive learning materials. Although the jury is still out on the debate about delayed versus immediate feedback in computer-based learning environments (Van der Kleij et al. 2015), the value of a computing device is that the learning task can be immediately complemented with directed feedback to best influence learning outcomes. We believe that the feedback messages should be immediate and not necessarily include response correctness, as providing the target verbal information (in a short message) appears to have a substantial effect on the learning outcomes irrespective of modality condition. Further, requiring the learner to assess the certitude of their own response can also have positive effects on the learning outcomes (Kealy and Ritzhaupt 2010).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27,* 669–681.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*(2), 416.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational psychology review, 14*(1), 5–26.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*(2), 75–100.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and format of instruction. *Cognition and Instruction, 8*(4), 293–332.

Cheon, J., Crooks, S., Inan, F., Flores, R., & Ari, F. (2011). Exploring the instructional conditions for a reverse modality effect in multimedia instruction. *Journal of Educational Multimedia and Hypermedia, 20*(2), 117–133.

Cohen, V. B. (1985). A reexamination of feedback in computer-based instruction: Implications for instructional design. *Educational Technology, 25*(1), 33–37.

Dwyer, F. M. (1965). An experimental evaluation of the relative effectiveness of selected visual illustrations in teaching science concepts to college freshmen. Unpublished Dissertation. The Pennsylvania State University, University Park, PA.

Dwyer, F. M. (1978). *Strategies for improving visual learning*. Learning Services, Box 784, State College, PA.

Dwyer, F. M., & Lamberski, R. (1983). The instructional effect of coding (color and black and white) in facilitating students' information acquisition and retrieval. *Educational Communication & Technology Journal, 31,* 9–21.

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction, 28,* 48–63.

Engle, R. W., & Roberts, J. S. (1982). How long does the modality effect persist? *Bulletin of the Psychonomic Society, 19*(6), 343–346.

Fiorella, L., Vogel-Walcutt, J., & Schatz, S. (2012). Applying the modality principle to real-time feedback and the acquisition of higher-order cognitive skills. *Educational Technology Research and Development, 60*(2), 223–238.

Gagne, R. (1985). *The conditions of learning* (4th ed.). New York: Holt, Rinehart & Winston.

Gagne, R., Briggs, L., & Wager, W. (1992). *Principles of instructional design* (4th ed.). Fort Worth, TX: HBJ College Publishers.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Abingdon: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81–112.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*(6), 717–742.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210–228.

Herrlinger, S., Höffler, T. N., Opfermann, M., & Leutner, D. (2016). When do pictures help learning from expository text? Multimedia and modality effects in primary schools. *Research in Science Education*. https://doi.org/10.1007/s11165-016-9525-y.

Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology, 23*(9), 1215–1226.

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146–154.

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18*(2), 178.

Johnson, C, & Priest, H. (2014). The feedback principle in multimedia learning. In *The Cambridge handbook of multimedia learning* (pp. 449–463). 2nd Edition. Cambridge: Cambridge University Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354.

Kealy, W. A., & Ritzhaupt, A. D. (2010). Assessment certitude as a feedback strategy for learners' constructed responses. *Journal of Educational Computing Research, 43*(1), 25–45.

Kollöffel, B., & de Jong, T. (2015). Can performance feedback during instruction boost knowledge acquisition? Contrasting criterion-based and social comparison feedback. *Interactive Learning Environments.* https://doi.org/10.1080/10494820.2015.1016535.

Levin, J. R. (1981). On functions of pictures in prose. In F. J. Pirozzolo & M. C. Wittrock (Eds.), *Neuropsychological and cognitive processes in reading* (pp. 203–228). New York: Academic Press.

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. *The Psychology of Illustration, 1,* 51–85.

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education, 60*(1), 95–109.

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

Mayer, R. E. (2009). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167–171.

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*(4), 444.

Mayer, R., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and instruction, 12*(1), 107–119.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149–158.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358.

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development, 53*(3), 35–45.

Morrison, G. R., Ross, S. M., Kemp, J. E., & Kalman, H. (2010). *Designing effective instruction*. Hoboken, NJ: Wiley.

Mory, E. H. (2004). Feedback research revisited. *Handbook of Research on Educational Communications and Technology, 2,* 745–783.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. *Handbook of Research on Educational Communications and Technology, 3,* 125–144.

O'Neil, H. F., Chuang, S. H. S., & Baker, E. L. (2010). Computer-based feedback for computer-based collaborative problem solving. In *Computer-based diagnostics and systematic analysis of knowledge* (pp. 261–279). New York: Springer.

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*(1), 110–117.

Paas, F., & Ayres, P. (2014). Cognitive load theory: A broader view on the role of memory in learning and education. *Educational Psychology Review, 26*(2), 191–195.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Journal of Psychology, 45*(3), 255–287.

Pastore, R. S. (2010). The effects of diagrams and time-compressed instruction on learning and learners' perceptions of cognitive load. *Educational Technology Research and Development, 58*(5), 485–505.

Pastore, R. (2012). The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Computers & Education, 58*(1), 641–651.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17*(4), 398–422.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Ritzhaupt, A. D., & Barron, A. (2008). Effects of time-compressed narration and representational adjunct images on cued-recall, content recognition, and learner satisfaction. *Journal of Educational Computing Research, 39*(2), 161–184.

Ritzhaupt, A. D., Barron, A. E., & Kealy, W. A. (2011). Conjoint processing of time-compressed narration in multimedia instruction: The effects on recall, but not recognition. *Journal of Educational Computing Research, 44*(2), 203–217.

Ritzhaupt, A. D., & Kealy, W. A. (2015). On the utility of pictorial feedback in computer-based learning environments. *Computers in Human Behavior, 48,* 525–534.

Schallert, D. L. (1980). The role of illustrations in reading comprehension. In *Theoretical issues in reading comprehension* (pp. 503–524).

Scheiter, K., Gerjets, P., & Heise, E. (2014). Distraction during learning with hypermedia: Difficult tasks help to keep task goals on track. *Frontiers in Psychology, 5,* 268.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction, 20*(2), 100–110.

Schneps, M. H., Thomson, J. M., Sonnert, G., Pomplun, M., Chen, C., & Heffner-Wong, A. (2013). Shorter lines facilitate reading in those who struggle. *PLoS ONE.* https://doi.org/10.1371/journal.pone.0071161.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Stevens, J. (1990). *Intermediate statistics: A modern approach.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York, NY: Springer.

Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes A meta-analysis. *Review of Educational Research, 85*(4), 475–511.

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95–99.

Wager, W., & Wager, S. (1985). Presenting questions, processing responses, and providing feedback in CAI. *Journal of Instructional Development, 8*(4), 2–8.

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71,* 79–89.

Albert D. Ritzhaupt is an associate professor at the University of Florida. His primary research areas focus on the design and development of technology-enhanced learning environments, and technology integration in education. His publications have appeared in multiple venues, including the Journal of Research on Technology in Education, Computers in the Schools, Journal of Computing in Higher Education, Computers & Education, Behavior Research Methods, Journal of Interactive Learning Research, Educational Technology Research and Development, Journal of Educational Computing Research, and Computers in Human Behavior.

Ray Pastore has multiple years of instructional design experience, which includes extensive corporate, K-12, and higher education experience. He earned his Ph.D. in Instructional Systems with a minor in Educational Psychology from Penn State University in 2008 and is currently an Associate Professor and Program Coordinator of the Instructional Technology master's program at the University of North Carolina Wilmington.

Jiahui Wang currently is a third-year Ph.D. student and research fellow in Educational Technology in College of Education at the University of Florida. She graduated from the University of Virginia with a M.Ed. Degree in Curriculum & Instruction and she taught Chinese at elementary and secondary levels in Fairfax County Public School, VA after graduation. Her future goal is to carry out interdisciplinary research that can provide valuable insights on the design and development of technology-rich learning environments.

Robert O. Davis is an assistant professor in the English Literature and Language Technology Department at Hankuk University of Foreign Studies in Seoul, South Korea. His current research interests involve pedagogical agent gesturing, social acceptance of virtual characters, interaction with computer-based environments, and virtual reality in the foreign language classroom.