**International Journal of Mobile and Blended Learning**
Volume 14 • Issue 1

# Images and Text on PowerPoint Slides: "Tracking" Their Impact on Information Retention

Joshua L. Williams, Georgia Southern University, USA*
Nancy McCarley, Georgia Southern University, USA
Stephen D. Willis, Georgia Southern University, USA
Amber N. Huddleson, Georgia Southern University, USA

## ABSTRACT

In this study, the authors examined the impact of irrelevant visual images in a PowerPoint lecture on attention and information retention performance. They found students viewing text-only PowerPoint slides retained less information than students viewing text-and-image slides, but the difference did not reach significance. However, when examining both fixation counts and fixation duration, they found students viewing text-and-image PowerPoint slides who spent more time looking at the text retained more information. They discuss the impact of varied visual attention (moving back and forth from text to image) on information integration and retention and establish empirical questions for future research.

## KEYWORDS

Eye-Tracking, Instructional Technology, Multimedia Learning

## INTRODUCTION

In the last 20 years, pedagogical techniques in higher education classrooms have grown and developed tremendously. A major catalyst for such growth and development has been the advent of novel technology. PowerPoint™ emerged as a presentation tool in the late 1980s and continues to be one of the key teaching tools in academia, both in and out of the classroom. In contemporary higher education, it seems that use of such presentation tools has grown exponentially as more and more classes are being offered in fully online, hybrid, live Zoom-delivered lectures, synchronous instruction, or flipped-classroom formats, with students spending significant time viewing pre-recorded lectures that include PowerPoint presentations with instructor voice over (Becker et al., 2018; Goodwin & Miller, 2013; Strelan et al., 2020). Further, the Covid-19 pandemic forced the use of more technology at the university level in order to continue offering courses (Daniel, 2020; Guraya, 2020; Rapanta et al., 2020). And, it is quite likely that even after we move past the pandemic, higher education administrators and instructors will have recognized the positive aspects of the many technological advances to support learning, which will drive more integration of such technology into university classes (Goh & Sandars, 2020). So, given that such presentation technology is even more integral in contemporary higher education, and likely will be for the foreseeable future, it is important to continue working to understand how students learn best in situations where such technology is used.

There is a history of mixed results related to the effectiveness of lectures accompanied by PowerPoint to enhance student motivation, note taking, interest, information retention and transfer,

DOI: 10.4018/IJMBL.291979
*Corresponding Author

among other things (Apperson et al., 2006; Lowry, 1999; Mantei, 2000; Susskind, 2005). One aspect of the mixed results may simply be due to the complex milieu of behaviors necessary for the learner to be successful in a dynamic multimedia presentation context. They must visually scan the PowerPoint slide, attend to relevant text, images, and animations; listen to the instructor as they speak about the content portrayed on the slide as well as other relevant spoken information; attempt to develop coherent connections among relevant information presented (both visually and orally); and finally integrate the incoming information with existing knowledge. And, if that is not sufficiently complex, the learner must do this in a more-or-less continuous, nonlinear, and parallel fashion given that the presentation will progress through different topics and slides for an extended period of time (Mayer, 2005; Williams et al., 2017). Further, another potential determinant of the mixed results could be due to the interaction of additional factors, such as the instructor's enthusiasm and skill at using the technology, the student's ability to attend to relevant information during the presentation and get it into their notes (if they are taking notes), the student's interest in the topic, and even the text, images, and animations used to structure the PowerPoint (Apperson et al., 2006; Clark, 2008; Gier & Kreiner, 2009; Nouri & Shahid, 2005; Williams et al., 2017). Mayer (2005) stated that how instructors actually structure their PowerPoint slides reveals how they believe a student learns in a multimedia context. With that statement in mind, he recommends that designers of the multimedia presentation, in order to help the learner be successful, should structure the presentation in such a way as to capitalize on how the mind functions; specifically using empirical work from cognitive psychology to inform design practices (Mayer, 2005).

## Cognitive Theory of Multimedia Learning

Mayer and colleagues (Mayer, 2001, 2005, 2008, 2017; Mayer et al., 2001; Moreno & Mayer, 2002), over the last several decades, have used the Cognitive Theory of Multimedia Learning (CTML) to drive a research program geared toward better understanding the mechanisms that underlie multimedia learning. This research team developed a body of work that provides an empirically based understanding of how humans learn best within a multimedia context by capitalizing on what we know from cognitive research. Rooted in theoretical and empirical work (Baddeley, 1999; Chandler & Sweller, 1991; Clark & Paivio, 1991; Wittrock, 1990), the CTML operates on three key assumptions: a) Humans have separate processing channels for visual and auditory information, b) Humans can only process a limited amount of information at once, and c) Humans are active, not passive, perceivers and processors of information. Research operating on these key assumptions led to many pragmatic suggestions for instructors to improve the design of their presentations to align with how students learn best in a multimedia environment (Mayer, 2008, 2017).

One key suggestion that emerged from this line of research was for instructors to use PowerPoint slides that contained only images relevant to the concurrent verbal presentation rather than PowerPoint slides that contained only relevant text. Or, at the very least, replace large amounts of text on the slides with images (Horvath, 2014). Such proposals came from work that revealed concurrent activation of the same perceptual networks when reading words, such as those on a PowerPoint slide, and listening to spoken words, resulting in increased cognitive load via the use of the same processing channel (Savoy et al., 2009). Such a situation is known as the verbal redundancy principle (Mayer, 2001; Mayer et al., 2001; Moreno & Mayer, 2002) and demonstrates the inverse relationship between cognitive load and learning. Since the learner is presented with two concurrent streams of information, they either selectively focus on one stream of information (either the words on the slide or the words being spoken) while processing and comprehension of the other suffers. Or, they attempt to divide their attention to focus on both and, as a result, processing and comprehension of both suffer (Yue et al., 2013; Wecker, 2012). The proposal to remove text, either fully or partially, should help reduce the cognitive load because text and images are not processed via the same channels (Horvath, 2014). Indeed, researchers have discovered better recall performance in conditions with image-only slides relative to both text-only and text-and-image combined PowerPoint slides (Hallett & Faria, 2006; Jamet & Le Bohec, 2007).

However, in reality, instructors design PowerPoint presentations in a plethora of ways, some of which do not involve any instructor design at all, but rather adoption of textbook-provided slides which may only have text-based slides. However, instructors who do follow the research may strike a compromise between the work in the CTML tradition (Mayer, 2001, 2005) and other work that revealed students' preferences and interest for not only PowerPoint, but specific structural components of the PowerPoint such as bulleted topical text phrases, images and graphs, colored backgrounds, and sounds to support the images (Apperson et al., 2006). Thus, in such a compromise, instructors may opt for their PowerPoint slides to contain some images and some text.

## Eye-Tracking Technology in Multimedia Learning

In recent years, eye-tracking technology has allowed researchers to begin understanding visual attention behavior across a variety of contexts, including during PowerPoint presentations (Shayan et al., 2017; Slykhuis et al., 2005). With such technology, researchers can collect precise measures to indicate location and duration of visual fixations, length and number of eye movements, amount of visual switching between locations, as well as other specific measures to examine emergent patterns in looking behavior. One relevant question to be explored with such technology is, "how does PowerPoint structure impact visual attention?" Slykhuis et al. (2005) used eye-tracking to examine how a small group of preservice teachers visually attended image and text zones on prerecorded PowerPoint slides with and without oral narration. They also examined visual attention to slide images as a function of image classification (Pozzer & Roth, 2003). Within this image classification, they had decorative (no caption or reference to them in slide text) and complementary (had captions and directly referenced in slide text) images. They discovered that participants spent significantly more time visually attending complementary images than decorative ones, a pattern which also seemed to hold in the presence and absence of narration. One other result to note is that when examining time spent in the text zone versus the image zone, participants spent a greater amount of time overall visually examining the text zone with decorative images present, regardless of narration level whereas such a pattern only emerged with complementary images when narration was absent (Slykhuis et al., 2005).

Although Slykhuis et al. (2005) demonstrated that looking behavior varied as a function of PowerPoint structure (images, text, and narration), one key aspect they did not address was how such visual attention behavior related to retention of information presented during the presentation. In 2011, Tangen and colleagues examined the impact of PowerPoint structure on retention of presentation information. In their study, Tangen et al. (2011) had three PowerPoint conditions: a) Slides with congruent images (one image on slide related to narration), b) Slides with incongruent images (one image on slide loosely related to narration), and c) Slides with text only (two or three bullet points per slide related to narration). As participants viewed the presentation, they provided ratings of their level of interest and then after viewing the recorded presentations, each participant completed a test over the presented information. Tangen et al. (2011) discovered participants' interest levels to be higher in both image conditions relative to the text-only condition. Despite interest being higher in both image-based conditions, participants performed significantly worse in the incongruent image condition than in the congruent image and text-only conditions. There was no difference in performance between the congruent image and text-only conditions, thus not supporting the verbal redundancy principle (text-only with narration should have performed worse; Mayer et al., 2001).

## Current Study

These two studies, in particular, revealed looking behavior (Slykhuis et al., 2005) and information retention (Tangen et al., 2001) varied as a function of PowerPoint structure (images, text, narration), but how actual looking behavior specifically related to such variation in retention performance remains an open question. In the current study, we built on this work in such a way as to combine visual attention and information retention measures during a PowerPoint presentation constructed in an ecologically valid way. We used remote eye-tracking technology as college students viewed 1 of 2 pre-recorded multimedia PowerPoint presentations on the topic of personality psychology: a) Text-only slides with narration or b) Text-and-image slides with narration. Given that instructors are likely to have both images and text on PowerPoint slides (rather than one or the other as is the case in more sterile laboratory studies), we opted for a more realistic condition to have text (bullet points relevant to narration) paired with images (irrelevant to narration; similar to decorative images in Slykhuis et al., 2005 and incongruent images in Tangen et al., 2011). After viewing the presentation, all participants completed a short quiz over the material covered in the presentation.

Given the theoretical and empirical work on the verbal redundancy principle (Mayer et al., 2001) and students' preference for and interest in PowerPoint presentations with images (Apperson et al., 2006; Tangen et al., 2011), we expected students who viewed the text-only presentation to demonstrate lower overall performance on the content quiz. However, for students in the text-and-image condition, we expected students who spent a greater amount of time visually attending the images to show lower quiz performance.

## METHOD

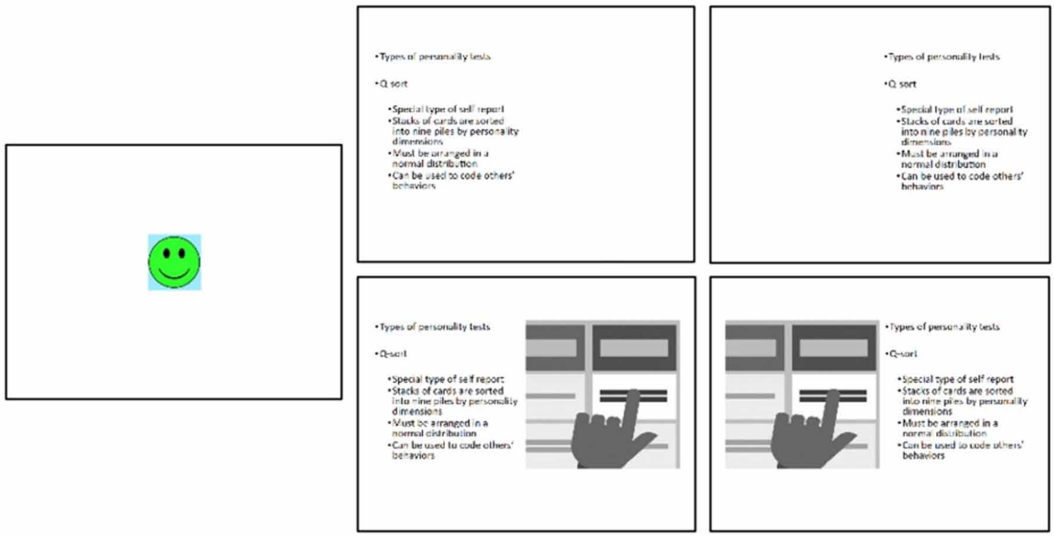
### Participants

After Institutional Review Board approval, 60 undergraduate students (39 women, 19 men, 2 preferred not to answer; $M_{age}$ = 24.38 years $SD_{age}$ = 10.23) at a university in the Southeastern United States participated and earned credit toward a research participation module in their introductory psychology course. In this sample, 33.9% ($n$ = 20) were First Year students, 27.1% ($n$ = 16) were Sophomores, 15.3% ($n$ = 9) were Juniors, and 23.7% ($n$ = 14) were Seniors (one participant preferred not to answer). Further, individuals comprising the sample self-identified as 1.7% ($n$ = 1) American Indian or Alaska Native, 5.0% ($n$ = 3) Asian, 15.0% ($n$ = 9) Black or African American, 8.3% ($n$ = 5) Hispanic, Latino, or Spanish Origin, 1.7% ($n$ = 1) Middle Eastern or North African, 61.7% ($n$ = 37) White, and 6.7% ($n$ = 4) Other (2 Hispanic and White; 1 American Indian, Asian, and White; 1 Middle Eastern or North African and White).

### Materials and Procedure

All participants visited our laboratory on campus to participate in a face-to-face and individualized manner (one participant at a time). We randomly assigned participants to 1 of 2 PowerPoint slide conditions: a) Text-only or b) Text-and-image. We used Camtasia 9 (Techsmith) to create an .mp4 video of the PowerPoint presentations with instructor voice over. The instructor, narration, and presentation length (14 minutes) were the same regardless of slide condition and each PowerPoint presentation covered the same topic in personality psychology (types of personality tests). We selected this topic because students in Introduction to Psychology had not been exposed to it at that point in the semester. Regardless of image presence, the text was the same on the slides and structured in such a way that the number of bullet points on each slide were equal (two heading bullets, four informational bullets), with the content for each bullet point relevant to the instructor's narration. Images used were basic, freely available black-and-white images that were not captioned nor referenced in the instructor's narration (decorative images as in Slykhuis et al., 2005; incongruent as in Tangen et al., 2011). To control for any directional looking bias, we counterbalanced the location of the text and images such that the specific location conditions were randomized across participants: Text left, text right, text left with image right, text right with image left. In addition, at each slide transition we ensured that all participants' visual attention started in the center of the slide by flashing a smiley face image briefly (500 *ms*) in the center prior to the appearance of the text/image, see *Figure 1* for exemplar PowerPoint stimuli.

**Figure 1. Exemplar PowerPoint stimuli**




Far left: Slide with smiley face that flashed briefly (500 *ms*) to capture attention to center of screen prior to appearance of text or text-and-image. Top two panels: Text-only stimuli (left and right versions for counterbalancing). Bottom two panels: Text-and-image stimuli (left and right versions for counterbalancing). All images used in the text-and-image stimuli were freely available, black-and-white images.

We collected all visual attention data with a RED250 eye tracking system (SensoMotoric Instruments [SMI], Inc.), which used a remote, binocular infrared camera to capture eye movements at 60 Hz. Participants sat individually in front of a 15.6" laptop, on which the tracking camera collected eye data from the bottom of the screen. We presented all aspects of the experiment using SMI's Experiment Suite, such that all participants engaged in an eye-tracker calibration and validation, viewed their randomly-assigned presentation, and completed a quiz over the presentation material within the software. Specifically, participants engaged in a 5-point visual calibration with a 4-point validation of the eye-tracking system (mean visual angle error = 0.43). Once calibration was complete, the experimenter started the presentation and left the testing room. The testing room was a low-distraction room, which contained no other objects except the laptop and there were no pictures on the walls. Further, during the presentation, the experimenter observed from an adjacent room via a one-way mirror. Once the participant viewed the presentation, they answered 21 multiple choice questions about the presentation content on the screen. Once they answered the questions, the experimenter returned to the testing room, had participants complete a brief demographic questionnaire, and then debriefed the participants.

## MEASURES AND ANALYSIS

We used SMI's BeGaze software to analyze eye movement data. On each slide, we created two areas of interest (AOI): one that covered the image (or space where the image would be on text-only slides) and another that covered the text. Each AOI on the slides was 218,202 pixels in size and each one covered 20.9% of the total space on the calibrated screen space (the PowerPoint slide in slideshow mode). The software computed the number of fixations per second, fixation duration (*ms*), and revisits per second (a measure akin to recurrence of behavior) to each AOI. Use of statistical procedures depended on whether data met assumptions of parametric tests. Specifically, we diagnosed data by consulting histograms and Shapiro-Wilk tests to determine if data conformed to normal distributions. When data did not meet the normality assumption, we used nonparametric procedures.

## RESULTS

### Manipulation Check

In the text-only condition, to determine if participants visually attended the area of the slide where an image was present in the text-image condition, we maintained the same area of interest (AOI) and compared the fixation count per second and fixation duration between the text-only and text-image conditions. As expected, participants in the text-image condition, in which an image was present on the slides, visually attended to the image AOI more than those in the text-only condition. Specifically, fixation count per second toward the image was significantly higher in the text-image condition ($M = 0.24$, $SD = 0.13$, Mean Rank = 45.10) than in the text-only condition ($M = 0.02$, $SD = 0.01$, Mean Rank = 15.90, $U = 12.00$, $z = -6.476$, $p < .001$, $r = .84$). Also, fixation duration to the image was significantly higher in the text-image condition ($M = 0.11$, $SD = 0.08$, Mean Rank = 45.07) than in the text-only condition ($M = 0.01$, $SD = 0.01$, Mean Rank = 15.93, $U = 13.00$, $z = -6.462$, $p < .001$, $r = .83$).

### Overall Quiz Performance

Although participants in the text-image condition ($M = 72.86$, $SD = 14.38$) scored higher on average than the text-only condition ($M = 66.67$, $SD = 18.04$), there was no significant difference in scores between the two conditions, $t\ (58) = 1.470$, $p = .147$, $d = 0.38$.

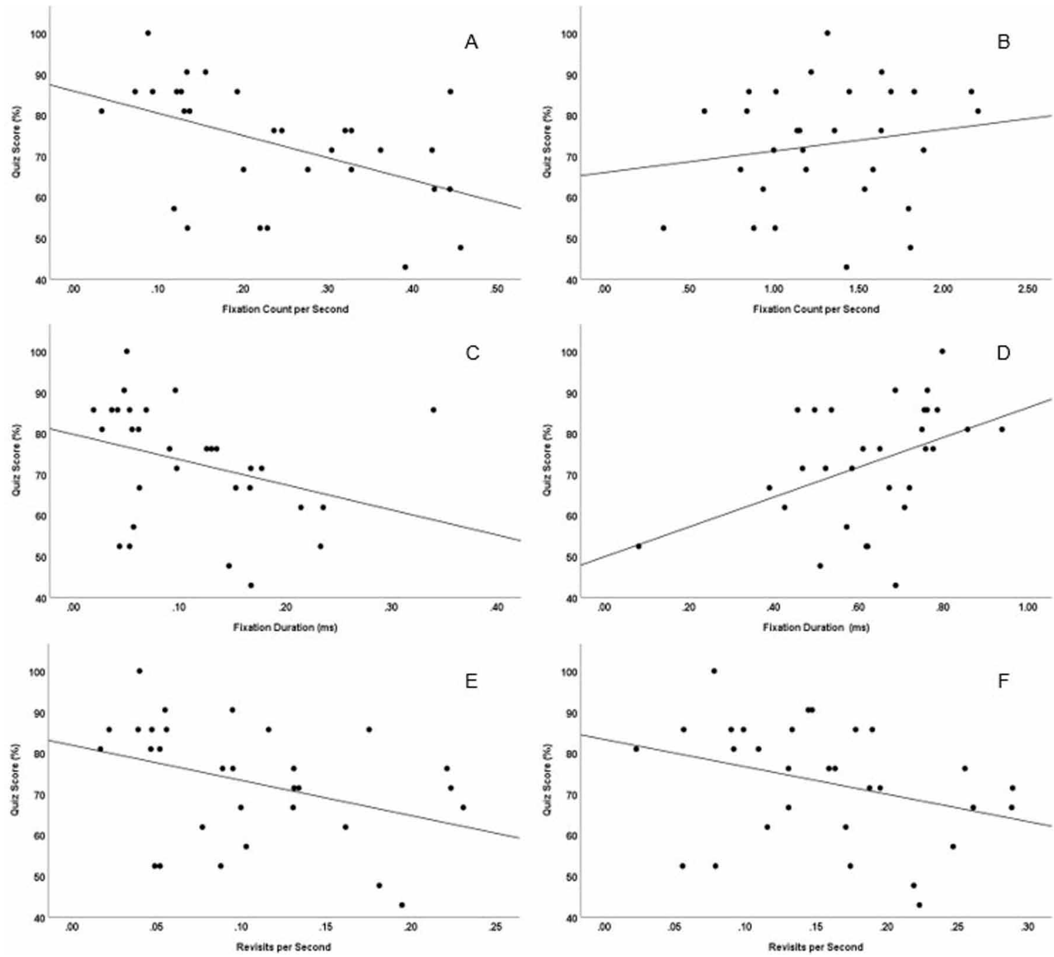
### Relationships Between Visual Measures and Quiz Performance

*Figures 2* and *3* display the relationship between fixations per second, fixation duration, and revisits per second on the image and text with quiz scores within the text-and-image and text-only conditions, respectively.

For the text-and-image condition, Spearman's rank-order correlation analyses revealed a significant negative monotonic relationship between fixation count per second toward images and scores on the quiz, $\rho(30) = -.48$, $p = .007$ and a positive, but non-significant relationship between fixation count per second toward text and scores on the quiz $\rho(30) = .17$, $p = .382$. There was a marginally significant negative monotonic relationship between fixation duration toward images and scores on the quiz, $\rho(30) = -.33$, $p = .077$ and a significant positive monotonic relationship between fixation duration toward text and scores on the quiz, $\rho(30) = .43$, $p = .017$. Revisits per second toward images had a significant negative monotonic relationship with scores on the quiz, $\rho(30) = -.37$, $p = .042$. Revisits per second toward text showed a marginally significant negative monotonic relationship with scores on the quiz, $\rho(30) = -.33$, $p = .076$, see *Figure 2*.

For the text-only condition, Spearman's rank-order correlation analyses revealed a negative non-significant relationship between fixation count per second toward the image area and scores on the quiz, $\rho(30) = -.22$, $p = .235$ and a marginally significant negative monotonic relationship between fixation count per second toward text and scores on the quiz $\rho(30) = -.33$, $p = .072$. There was a negative non-significant relationship between fixation duration toward the image area and scores on the quiz, $\rho(30) = -.20$, $p = .291$ and a positive non-significant relationship between fixation duration toward text and scores on the quiz, $\rho(30) = .28$, $p = .14$. Revisits per second toward the image area had a negative non-significant relationship with scores on the quiz, $\rho(30) = -.26$, $p = .165$. Revisits per second toward text showed a marginally significant negative monotonic relationship with scores on the quiz, $\rho(30) = -.32$, $p = .086$, see *Figure 3*.

**Figure 2. Scatterplots of fixations per second (A and B), fixation duration (C and D), and revisits (E and F) to the image (left panels) and text (right panels) for the text-image condition**

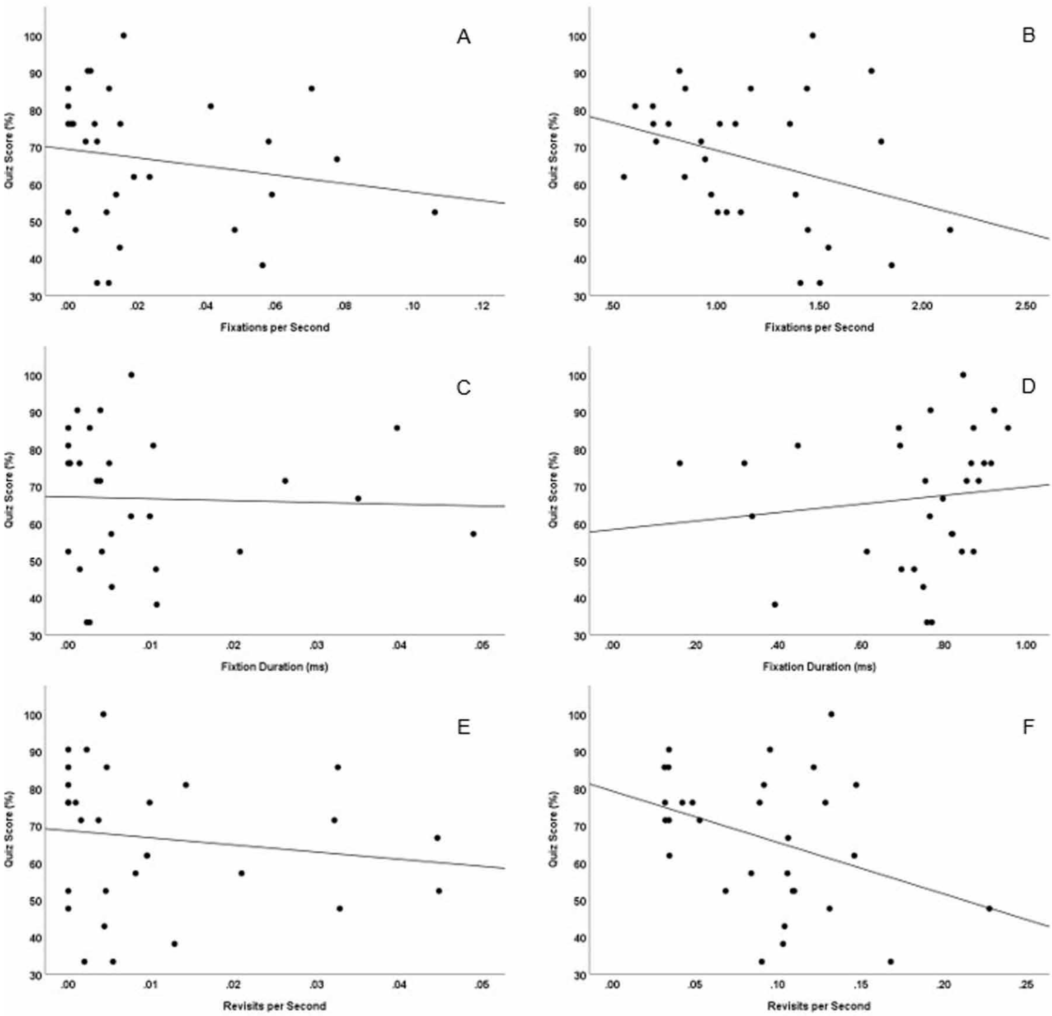



## Text-Image Condition: High-Low Visual Behavior and Quiz Performance

Given the consistent relationships found, when examining the text-and-image condition, we explored the impact of varied looking patterns to images and text when both were present on quiz performance. To examine differential looking patterns within the text-and-image condition, we used a median split on fixations per second, fixation duration, and revisits per second. This allowed us to conduct 2 (Image: High v. Low) x 2 (Text: High v. Low) factorial ANOVA with quiz score as the dependent variable. With this approach, the individuals that comprised the groupings for each factorial ANOVA slightly differed based on the eye-tracking metric used. See *Table 1* for a basic breakdown of descriptive information of the individuals who comprised each factor level. *Table 2* displays the cell means for the factorial ANOVA analyses for fixations per second, fixation duration, and revisits per second.

***Fixations per second.*** The median fixations per second toward the image was 0.22 and the median fixations per second toward the text was 1.27. There was a main effect of Image with those who had lower fixations per second on the image scoring higher than those who had higher fixations per second on the image, $F(1, 26) = 5.595$, $p = .026$, $\eta_p^2 = .18$. There was no main effect of Text on scores, $F(1, 26) = 0.012$, $p = .915$, $\eta_p^2 = .01$. There was no significant interaction between Image and Text, $F(1, 26) = 2.431$, $p = .131$, $\eta_p^2 = .09$, see *Table 2* and *Figure 4*.

**Figure 3. Scatterplots of fixations per second (A and B), fixation duration (C and D), and revisits (E and F) to the image (left panels) and text (right panels) for the text-only condition**




***Fixation duration.*** The median fixation duration toward the image was 0.09 and the median fixation duration toward the text was 0.66. There was no main effect of Image or Text on scores, $F(1, 26) = 2.036$, $p = .165$, $\eta_p^2 = .07$, $F(1, 26) = 2.618$, $p = .118$, $\eta_p^2 = .09$, respectively. There was a marginal interaction between Image and Text, $F(1, 26) = 2.934$, $p = .099$, $\eta_p^2 = .10$, see *Table 2* and *Figure 5*. Within those who had low fixation durations to the text, there was no difference in scores between those who had high or low fixation durations to the image ($p = .841$). Within those who had high fixation durations to the text, those who had low fixation durations to the image showed significantly higher scores than those who had high fixation durations to the image ($p = .035$). Within those who had low fixation durations to the image, those who had high fixation durations to the text showed significantly higher scores than those who had low fixation durations to the text ($p = .026$). Within those who had high fixation durations to the image, there was no difference in scores between those who had high or low fixation durations to the text ($p = .947$).

***Revisits per second.*** The median revisits per second toward the image was 0.09 and the median revisits per second toward the text was 0.15. There was no main effect of Image or Text on scores,

**Table 1. Basic descriptive information for high and low groupings in factorial analyses**

| | Mean Age (*SD*) | Sex | | College Level | | | |
|---|---|---|---|---|---|---|---|
| | | Female | Male | First Year | Sophomore | Junior | Senior |
| **Fixation Count (per *s*)** | | | | | | | |
| Text - High | 27.60 (*11.49*) | 10 | 3 | 5 | 2 | 3 | 5 |
| Text - Low | 21.07 (*3.65*) | 10 | 5 | 2 | 6 | 3 | 4 |
| Image - High | 22.13 (*4.63*) | 9 | 5 | 2 | 7 | 3 | 3 |
| Image - Low | 26.53 (*11.68*) | 11 | 3 | 5 | 1 | 3 | 6 |
| **Fixation Duration (*ms*)** | | | | | | | |
| Text - High | 24.13 (*6.92*) | 10 | 3 | 5 | 3 | 1 | 6 |
| Text - Low | 24.53 (*10.97*) | 10 | 5 | 2 | 5 | 5 | 3 |
| Image - High | 21.67 (*4.22*) | 10 | 4 | 2 | 6 | 3 | 4 |
| Image - Low | 27.00 (*11.63*) | 10 | 4 | 5 | 2 | 3 | 5 |
| **Revisits (per *s*)** | | | | | | | |
| Text - High | 23.13 (*5.29*) | 9 | 5 | 2 | 5 | 5 | 3 |
| Text - Low | 25.53 (*11.71*) | 11 | 3 | 5 | 3 | 1 | 6 |
| Image - High | 22.07 (*4.91*) | 9 | 4 | 3 | 5 | 3 | 3 |
| Image - Low | 26.31 (*11.28*) | 11 | 4 | 4 | 3 | 3 | 6 |

$F(1, 26) = 0.312$, $p = .581$, $\eta_p^2 = .01$, $F(1, 26) = 1.559$, $p = .223$, $\eta_p^2 = .06$, respectively. There was no significant interaction between Image and Text, $F(1, 26) = 0.035$, $p = .853$, $\eta_p^2 = .01$, see *Table 2*.

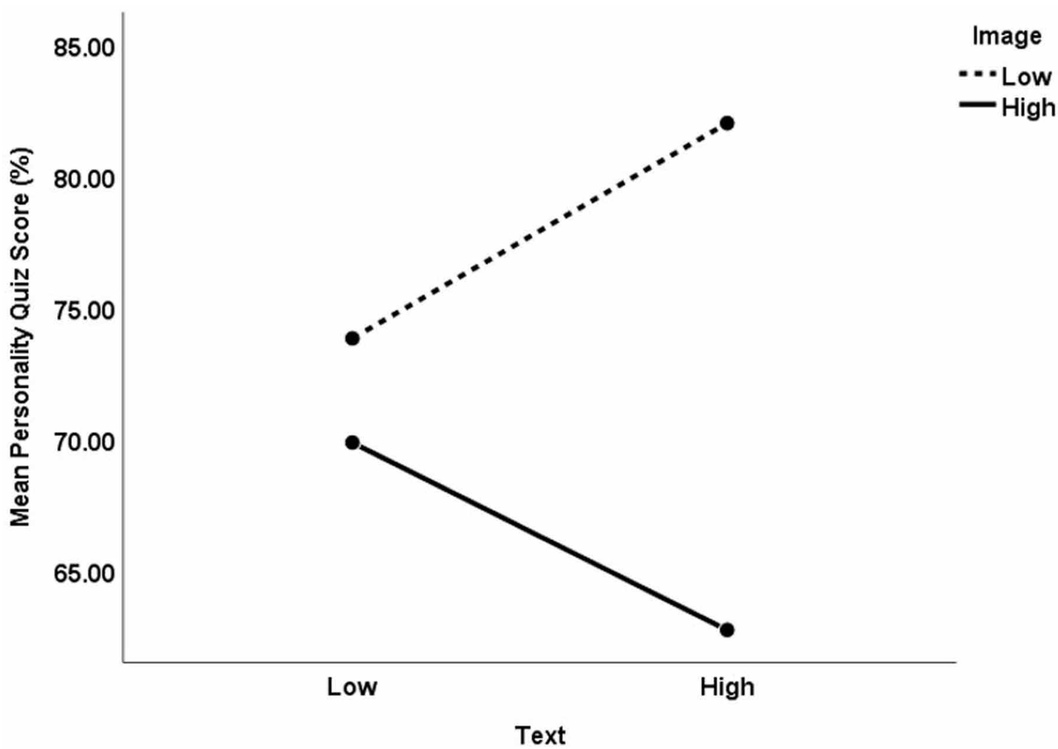
## DISCUSSION

In this study, we examined how text-only and text-and-image PowerPoint slides impacted visual attention patterns and quiz performance. Based on the verbal redundancy principle (Mayer, 2001; Mayer et al., 2001), we expected participants who viewed the text-only lecture with instructor narration

**Table 2. Cell means (SD) of quiz performance for fixation count, duration, and revisits**

| | Text – High | Text – Low |
|---|---|---|
| **Fixation Count (per *s*)** | | |
| Image – High | 62.70 (*14.57*) | 69.84 (*9.52*) |
| Image – Low | 82.01 (*12.77*) | 73.81 (*16.97*) |
| **Fixation Duration (*ms*)** | | |
| Image – High | 67.62 (*17.63*) | 68.10 (*11.46*) |
| Image - Low | 83.33 (*8.77*) | 66.66 (*17.49*) |
| **Revisits (per *s*)** | | |
| Image – High | 66.27 (*12.26*) | 76.19 (*13.46*) |
| Image - Low | 71.43 (*17.17*) | 78.75 (*14.63*) |

**Figure 4. Mean percent score on personality quiz for the 2 (Image fixations per second: Low v. High) x 2 (Text fixations per second: Low v. High) relationship.**

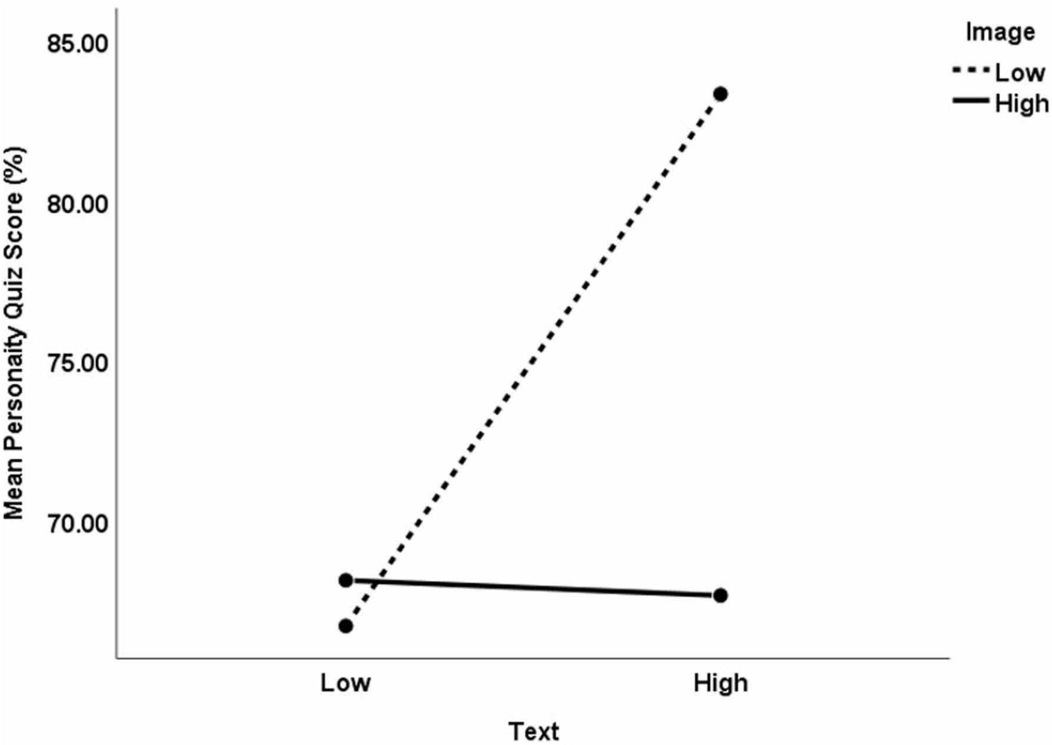



to perform worse on the quiz than participants who viewed the text-and-image presentation. Indeed, we discovered such a pattern in the results, but the difference did not reach significance.

When text and images were present concurrently on the slides, we also expected participants who spent more time visually attending the images, which were irrelevant to the narration [decorative (Slykhuis et al., 2005) and incongruent (Tangen et al., 2011)], to perform worse on the quiz. Such a pattern emerged when we viewed the relationships among fixations per second, fixation duration, and quiz performance. Specifically, as seen in *Figure 2* (panels A and C), as fixations per second and fixation duration toward the images increased, the quiz performance tended to decrease. Conversely, as fixations per second and fixation duration toward the bulleted text increased, the quiz performance tended to increase as well (*Figure 2*, panels B and D). Further, when we combined fixation duration toward images and text into the same analysis to examine how such behaviors may work together to impact quiz performance, we discovered an interaction pattern in line with our prediction. Specifically, we found a divergent interaction pattern in which participants who had low fixation durations toward images but high fixations toward text performed best on the quiz (see *Figure 5*).

The overall pattern of quiz performance was in line with the verbal redundancy principle (Mayer, 2001; Mayer et al., 2001) as participants in the text-and-image condition showed a higher mean score than participants in the text-only condition. Despite such a pattern, there was no statistically significant difference between the groups, a result consistent with the findings of Tangen et al. (2011). The lack of difference could certainly have been simply a product of a small sample size and lack of power to detect the effect. However, it could also be that there was simply no difference to detect. This could be due to the slide bullet points being designed in a more realistic way: shortened and not complete sentences. Thus, the cognitive load that may have been present due to the verbal redundancy principle

**Figure 5. Mean percent score on personality quiz for the 2 (Image fixation duration: Low v. High) x 2 (Text fixation duration: Low v. High) relationship**




may have been sufficiently lowered due to the participants not needing to engage in deeper processing to read long bullet points.

Consistent with the findings of Slykhuis et al. (2005), we found that varied PowerPoint structure impacted visual attention. In our study, we, too, found that participants spent a greater proportion of time visually attending the slide text relative to the irrelevant images (decorative as labeled by Slykhuis et al., 2005). However, here we were able to extend the work to relate the varied visual attention to quiz performance. Under the conditions of this study, we found that a greater amount of looking toward the irrelevant images (as measured by fixation count per second and fixation duration) was associated with lower performance on the quiz. It may be that spending more time visually examining the irrelevant images distracted participants from examining the narration-relevant text on the slides, or drove them to spend more time trying to discover connections between the irrelevant information in the image with the text and narration. Such an explanation may have some support in our revisits per second relationships in that as revisits to images and text increased, performance tended to decrease. In this case, it is possible that visually leaving and returning to the areas repeatedly may indicate participants trying to establish connections between the image and text, which do not actually exist. Although Tangen et al. (2011) did not have precise visual attention measures to know for sure, and we did not have a congruent image condition as in their study, it is possible that lower performance in their incongruent image condition may have been linked to a greater amount of time spent visually examining irrelevant information presented in the incongruent images and/or potentially trying to integrate it with the narration. In their congruent image condition, visual attention patterns to relevant information in the image may have facilitated integration of information with the narration.

However, this is not known from either study and provides a potential line of research to tease out such relationships.

### Limitations and Future Directions

This study was not without some key limitations. Given the goal of integrating research lines of PowerPoint structure and visual attention capture with eye-tracking, we sought to keep the study simple and test just a few specific hypotheses. As such, we did not have a text-and-image condition in which the images were relevant to the narration and slide text (like the complementary images of Slykhuis et al. (2005) and the congruent images of Tangen et al. (2011). The addition of a condition such as this will be important to determining if visual patterns and quiz performance vary between slides that have text with irrelevant images and slides that have text with relevant images. For instance, as noted earlier, would visual attention and attention-switching foster enhanced comprehension due to the relevance across text, image, and narration?

A second limitation of the current study may be a product of the high degree of experimental control in designing the PowerPoint slides with images. As in any sort of learning and perception study, there is a trade-off between needing to rule out alternative explanations, thereby enhancing internal validity, and trying to foster ecological validity. In many cases, PowerPoint lectures will contain colored images and complex graphics. In our case, we used bland, grayscale images to control for any impact color may have had in attracting attention. Indeed, color has been shown to impact visual attention in images (Frey et al., 2008). Future lines of research will examine relationships among color variations, visual attention, and presentation comprehension. However, despite lacking color in the images, we did create a PowerPoint presentation that was akin to one that would likely be similar to one that could be recorded and posted in a learning management system for online, hybrid, or flipped class structures, thereby carrying some ecological validity.

## CONCLUSION

In this study, we extended the work of Slykhuis et al. (2005) and Tangen et al. (2011) to combine precise visual attention measures with eye-tracking technology with controlled variations in PowerPoint structure to examine how such variations related to information retention performance. In doing so, we have established some key empirical questions for future research which will allow our laboratory, and others, to continue the work of Mayer and colleagues (Mayer, 2001; 2005; 2008; 2017; Mayer et al., 2001; Moreno & Mayer, 2002) in determining how students learn best from multimedia presentations.

### Conflicts of Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

### Funding Statement

No funding was received for this work.

### Process Dates:

Received: April 15, 2021, Revision: August 9, 2021, Accepted: August 19, 2021

### Corresponding Author:

Correspondence should be addressed to Joshua Williams; jlwilliams@georgiaSouthern.edu

## REFERENCES

Apperson, J. M., Laws, E. L., & Scepansky, J. A. (2006). The impact of presentation graphics on students' experience in the classroom. *Computers & Education*, *47*(1), 116–126. doi:10.1016/j.compedu.2004.09.003

Becker, S. A., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., & Diaz, V. (2018). *NMC Horizon Report: 2018 Higher Education Edition*. EDUCAUSE.

Clark, J. (2008). PowerPoint and pedagogy: Maintaining student interest in university lectures. *College Teaching*, *56*(1), 39–45. doi:10.3200/CTCH.56.1.39-46

Daniel, J. (2020). Education and the COVID-19 pandemic. *Prospects*, *49*(1-2), 91–96. doi:10.1007/s11125-020-09464-3 PMID:32313309

Frey, H., Honey, C., & König, P. (2008). What's color got to do with it? The influence of color on visual attention in different categories. *Journal of Vision (Charlottesville, Va.)*, *8*(14), 1–16. doi:10.1167/8.14.6 PMID:19146307

Gier, V. S., & Kreiner, D. S. (2009). Incorporating active learning with PowerPoint-based lectures using content-based questions. *Teaching of Psychology*, *36*(2), 134–139. doi:10.1080/00986280902739792

Goh, P., & Sandars, J. (2020). A vision of the use of technology in medical education after the COVID-19 pandemic. *Medpublish*, *9*(1), 49. doi:10.15694/mep.2020.000049.1

Goodwin, B., & Miller, K. (2013). Evidence on flipped classrooms is still coming in. *Educational Leadership*, *70*(6), 78–80.

Guraya, S. (2020). Combating the COVID-19 outbreak with a technology-driven e-flipped classroom model of educational transformation. *Journal of Taibah University Medical Sciences*, *15*(4), 253–254. doi:10.1016/j.jtumed.2020.07.006 PMID:32837509

Hallett, T. L., & Faria, G. (2006). Teaching with multimedia: Do bells and whistles help students learn? *Journal of Technology in Human Services*, *24*(2-3), 167–179. doi:10.1300/J017v24n02_10

Horvath, J. C. (2014). The neuroscience of PowerPoint. *Mind, Brain and Education: the Official Journal of the International Mind, Brain, and Education Society*, *8*(3), 137–143. doi:10.1111/mbe.12052

Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology*, *32*(4), 588–598. doi:10.1016/j.cedpsych.2006.07.001

Lowry, R. B. (1999). Electronic presentation of lectures – effect upon student performance. *University Chemistry Education*, *3*, 18–21.

Mantei, E. J. (2000). Using internet class notes and PowerPoint in physical geology lecture: Comparing the success of computer technology with traditional teaching techniques. *Journal of College Science Teaching*, *29*, 301–305.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge University Press. doi:10.1017/CBO9781139164603

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 31–48). Cambridge University Press. doi:10.1017/CBO9780511816819.004

Mayer, R. E. (2008, November). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *The American Psychologist*, *63*(8), 760–769. doi:10.1037/0003-066X.63.8.760 PMID:19014238

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403–423. doi:10.1111/jcal.12197

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187–198. doi:10.1037/0022-0663.93.1.187

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, *94*(1), 156–163. doi:10.1037/0022-0663.94.1.156

Nouri, H., & Shahid, A. (2005). The effect of PowerPoint presentations on student learning and attitude. *Global Perspectives on Accounting Education*, *2*, 53–73.

Pozzer, L. L., & Roth, W. M. (2003). Prevalence, function, and structure of photographs in high school biology textbooks. *Journal of Research in Science Teaching*, *40*(10), 1089–1114. doi:10.1002/tea.10122

Rapanta, C., Botturi, L., Goodyear, P., Guárdia, L., & Koole, M. (2020). Online university teaching during and after the Covid-19 crisis: Refocusing teacher presence and learning activity. *Postdigital Science and Education*, *2*(3), 923–945. doi:10.1007/s42438-020-00155-y

Shayan, S., Bakker, A., Abrahamson, D., Duijzer, C. A. C. G., & van der Schaff, M. (2017). Eye-Tracking the Emergence of Attentional Anchors in a Mathematics Learning Tablet Activity. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 166–194). IGI Global. doi:10.4018/978-1-5225-1005-5.ch009

Slykhuis, D. A., Wiebe, E. N., & Annetta, L. A. (2005). Eye-tracking students' attention to PowerPoint photographs in a science education setting. *Journal of Science Education and Technology*, *14*(5-6), 509–520. doi:10.1007/s10956-005-0225-z

Strelan, P., Osborn, A., & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review*, *30*, 1–22. doi:10.1016/j.edurev.2020.100314

Susskind, J. E. (2005). PowerPoint's power in the classroom: Enhancing students' self-efficacy and attitudes. *Computers & Education*, *45*(2), 203–215. doi:10.1016/j.compedu.2004.07.005

Tangen, J. M., Constable, M. D., Durrant, E., Teeter, C., Beston, B. R., & Kim, J. A. (2011). The role of interest and images in slideware presentations. *Computers & Education*, *56*(3), 865–872. doi:10.1016/j.compedu.2010.10.028

Wecker, C. (2012). Slide presentations as speech suppressors: When and why learners miss oral information. *Computers & Education*, *59*(2), 260–273. doi:10.1016/j.compedu.2012.01.013

Williams, J. L., McCarley, N., Sharpe, J. L., & Johnson, C. (2017). The ability to discern relevant from irrelevant information on PowerPoint slides: A key ingredient to the efficacy of performance feedback. *North American Journal of Psychology*, *19*(1), 219–236.

Wittrock, M. C. (1990). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376. doi:10.1207/s15326985ep2404_2

Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*, *105*(2), 266–277. doi:10.1037/a0031971

*Joshua L. Williams is Associate Chair and Associate Professor of Psychology at Georgia Southern University. He earned a B.A. in Health and Kinesiology from Purdue University and M.A. and Ph.D. in Psychology (Experimental-Developmental) from the University of Tennessee, Knoxville.*

*Nancy McCarley is an Associate Professor of Psychology at Georgia Southern University. She earned her B.S. and M.S. in Psychology and her Ph.D. in Educational Psychology from Mississippi State University.*

*Stephen D. Willis is a recent graduate of Georgia Southern University where he earned a B.S. in Psychology. Currently, Stephen is pursuing a career in behavior analysis and aspires to start graduate school with the goal of attaining a BCBA (Board Certified Behavior Analysis) certification.*

*Amber N. Huddleson is a post-baccalaureate pre-medical student at Georgia Southern University. She earned a B.S. in Psychology from Georgia Southern University.*