Received: 7 December 2017 | Revised: 25 September 2018 | Accepted: 25 September 2018
DOI: 10.1002/acp.3469

RESEARCH ARTICLE

WILEY

Student's metamemory knowledge about the impact of stereoscopic three-dimensional presentations of science content

John Dunlosky[1] | **Daniel Dudley**[1] | **Mary Beth Spitznagel**[1] | **Robert J. Clements**[2]

[1] Department of Psychological Sciences, Kent State University, Kent, Ohio, USA
[2] Department of Biological Sciences, Kent State University, Kent, Ohio, USA

Correspondence
John Dunlosky, Department of Psychological Sciences, Science of Learning and Education Center, Kent State University, Kent, OH.
Email: jdunlosk@kent.edu

Summary

We investigated students' knowledge and beliefs about the impact of using three-dimensional (3D) multimedia presentations. Students listened to a lecture about the ventricular system, which was presented alone (Experiment 1 only) or with a 3D or a 2D video illustrating the system. Afterwards, students judged how well they would perform on a criterion test. In Experiment 1, students judged that the 3D presentation would be superior to listening to the lecture alone (d = 0.81). Mean judgments were higher for the 3D than 2D presentation (d = 0.24), but this difference was not significant, so we estimated the effect size again. In Experiment 2, judgments were significantly higher after the 3D than 2D presentations (d = 0.40). Test performance was not significantly greater after the 3D than 2D presentations. A survey study again revealed that students believe 3D presentations are superior, and most students reported preferring them to 2D presentations.

KEYWORDS

3D presentations, education, metamemory, multimedia, stereoscopic presentations

1 | INTRODUCTION

Cognitive and educational scientists have been exploring innovative techniques to improve student learning for decades. A general approach involves discovering the best ways to present complex content to help students better understand it, and a specific approach that can improve student learning is to use multimedia presentations, such as by presenting causal diagrams along with lectures (for general guidelines for developing effective multimedia presentations, see Mayer, 2009, in press). Recent research investigating multimedia effects has demonstrated that supplementing lectures with stereoscopic three-dimensional (3D) presentations of spatially dense information can produce at least as good (and often better) performance as compared with using monoscopic two-dimensional (2D) presentations (Dongmei, Wilson, Rockhold, Lehman, & Lynch, 2017; Ferdig, Blank, Kratcoski, & Clements, 2015; Kaufmann, Schmalstieg, & Wagner, 2000; Perry, Cunningham, Gamage, & Kuehn, 2011; Petersson, Sinkvist, Wang, & Smedby, 2009; Wang, Chang, & Li, 2006).[1] For instance, the use of 3D presentations of some complex science concepts can enhance learning outcomes (for a comprehensive review, see McIntire, Havig, & Geiselman, 2014). As little as a decade ago, the promise of using 3D technology would have been less significant for improving education, because availability and equipment costs would have limited its use by only the most highly funded schools. However, the costs of this technology has dropped and the availability of cellular telephones, and the emergence of inexpensive virtual reality platforms capable of displaying 3D images is making its widespread use a possibility (for detailed discussion, see Brown & Green, 2016).

Given that educators may consider using this technology in the classroom, our primary goal was to extend prior research (which has

---

[1] In the present research, graphics were presented either stereoscopically (producing 3D depth of view from independent left and right eye views) or monoscopically (2D display with a single view for both eyes). For the sake of brevity, we will refer to stereoscopic and monoscopic presentations as 3D and 2D presentations, respectively.

largely focused on performance measures) by providing answers to the following questions: Do students believe that their learning is better after a multimedia presentation (with a lecture supplemented by a 3D or 2D video) than after hearing a lecture alone? And, more important, do students believe that their learning is better after multimedia presentations that include a 3D versus 2D presentation? Finally, do they prefer viewing 3D presentations over 2D ones in educational settings? Answering these questions is relevant to understanding one component of students' *metamemory*, which is a multifaceted concept that pertains to (a) one's knowledge about how memory operates, (b) one's monitoring of their memory, and (c) one's control of memory (for an extensive introduction to metamemory, see Dunlosky & Metcalfe, 2009). Our current focus is on students' metamemory *knowledge and beliefs* (the first component above) because they can influence the decisions students make about how to regulate their learning. For instance, a student who believes rereading is the best strategy to learn course materials while preparing for an exam would likely use this relatively ineffective strategy. Unfortunately, students hold inaccurate beliefs about the relative effectiveness of a variety of learning techniques (Bjork, Dunlosky, & Kornell, 2013), which motivates our central question: What beliefs do students have about the impact of 3D presentations on learning? To answer this question in the present research, we used two methods (a judgment method, Experiments 1 and 2, and a survey method) in an attempt to reveal students' knowledge and beliefs about stereoscopic 3D presentations.

Before we describe details of the present research, we first consider evidence from prior research that is relevant to students' metamemory knowledge about multimedia presentations. In this research, participants typically study a multimedia presentation (e.g., a text with static pictures) or a presentation with a single format (e.g., text alone), and then judge how well they have learned the target information in the text (e.g., Cardwell, Lindsay, Foerster, & Garry, 2017; Eitel, 2015; Ikeda, Kitagami, Takahashi, Hattori, & Ito, 2013; Serra & Dunlosky, 2010; Wiley, Sarmento, Griffin, & Hinze, 2017). The key outcome is whether judgments are higher for the multimedia presentation, which suggests the participants believe that the multimedia presentation is superior for learning. Consider outcomes from Serra and Dunlosky (2010), who had college students read a six-paragraph text that described how lightning storms develop. For some participants, the paragraphs were presented alone during study; for other participants, each paragraph was accompanied by an explanatory diagram. After reading the text, participants made a global judgment about how well they would perform on an upcoming test over the content. Across multiple experiments, the global judgments were higher for the multimedia presentation. Moreover, the multimedia presentation did not always boost actual test performance, so students judged multimedia presentations as superior even when they were inert (see also, Cardwell et al., 2017; Ikeda et al., 2013; for an exception, see Jaeger & Wiley, 2014). These results suggest the *multimedia-superiority hypothesis*, which is simply that people believe that adding pictorial information to a lecture generally benefits memory.

In the present research, the lecture described and labeled ventricles of the human ventricular system, and participants' objective was to learn the names of the ventricles so that they could correctly label them on a subsequent test. Although the script for the lecture (Appendix A) was written so that participants listening to only the lecture could potentially label the ventricles, doing so without the multimedia presentation was expected to be rather challenging. Accordingly, a prediction in the present context is that college students will judge multimedia presentations (either 3D or 2D) as more effective as compared with presenting the lecture alone. Note, however, prior research on students' judgments involving multimedia presentations have used static presentations and not dynamic multimedia presentations as used in the present research. Thus, a possibility is that students will find the dynamic video (described in Method section 2.1) distracting and hence will judge that it undermines their learning.

More important, will students judge that their learning is better when multimedia presentations include a 3D rather than a 2D presentation? Answers to this question are less clear because prior studies have not explored students' knowledge and beliefs about 3D versus 2D multimedia presentations. According to the cue-utilization framework of metamemory judgments (Koriat, 1997), people presumably do not have direct access to how well to-be-learned content has been encoded in memory, but instead, they infer their learning based on a variety of cues. The 3D versus 2D formats would produce different cues that could lead students to judge 3D (vs. 2D formats) as either more or less effective, depending on how the students interpret the relevance of those cues for their learning. On one hand, a reason that students may judge 3D presentations as more effective is that the to-be-learned content (ventricles in the brain) is three dimensional, and more of these cues (in terms of realistic depth cues and spatial layout) are included in a 3D than in a 2D presentation. On the other hand, some of the extra cues in 3D presentations may be viewed as distracting or less useful, and if so, then students may believe that 3D presentations are no better (and perhaps even worse) than are 2D presentations. The present research is the first to explore these competing predictions.

To test them, we used two methods to provide converging evidence for our conclusions. In the first two experiments, we had students judge their learning after lectures supplemented with 3D presentations, 2D presentations, or without the video presentation of the ventricles (the latter in Experiment 1 only). The lecture was about the ventricular system of the brain. Students listened to the lecture (either with or without the multimedia supplement) and then made a global prediction about their upcoming performance. In the final study, we had students first watch both the 3D and the 2D presentations, and then they completed a survey about their beliefs and preferences regarding them. Outcomes were relevant to evaluating the multimedia-superiority hypothesis (Experiment 1 only) as well as students' metamemory knowledge about the differential impact of 3D versus 2D multimedia presentations.

A secondary goal of the present research was to obtain an estimate of the benefit of 3D learning in this domain. Published reports have shown benefits to performance for visually complex materials after 3D than 2D presentations (for a review and discussion of exceptions, see McIntire et al., 2014). Accordingly, we chose the ventricular system as the target material, given that it is an anatomical system with an overlapping 3D structure that can be difficult to visualize. To foreshadow, the benefits were not consistently significant (although performance was always at least as high for 3D than for 2D), which allow us to evaluate whether students' judgments consistently followed the pattern of performance outcomes or instead were more likely being based on a general belief about the perceived benefit of 3D presentations.

2 | EXPERIMENT 1

2.1 | Method

2.1.1 | Design and participants

Presentation format (lecture alone, lecture and 2D presentation, and lecture and 3D presentation) was manipulated between participants. Ninety undergraduates (*Mean* age = 20, *SD* = 3.2; female percentage = 71%) from a large state university in the Midwest participated from a psychology participant pool to receive course credit. A power analysis conducted using G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007) with power set at 0.80 and two-tailed $\alpha = 0.05$ indicated that this sample size afforded sufficient sensitivity to detect between-group effects of $d \geq 0.74$. We expected students to have little to no knowledge about brain structure (as was later confirmed by the poor performance for the group hearing the lecture only), but importantly, we randomly assigned participants ($n = 30$) to each of the three groups. Performance on the paper folding task did not significantly differ between groups: Means (standard deviations, *SD*, in parentheses) were 4.1 (2.3), 4.3 (1.7), and 5.5 (2.1), for the lecture-alone, 2D, and 3D presentations, respectively, $F(2, 87) = 3.91$, MSE = 4.27. Given these unexpected differences, we used performance on the paper folding task as a covariate when conducting analysis of variance for Experiment 1; note, however, that the same statistical conclusions were supported regardless of whether performance on this task was included as a covariate.

2.1.2 | Materials

A lecture that named and described the relation among the ventricles of the brain was recorded by the first author. The lecture was approximately 2 min and 40 s long, and the transcript of the lecture is presented in Appendix A. The video presented to participants differed by group. For the lecture alone group, the names of the ventricles were presented on the screen with the same timing (onset and offset), placement on the screen, and format as those names that appeared in the 2D and 3D videos. Thus, participants in the lecture only group could read the ventricle names as they were spoken during the lecture. As noted in the introduction, the script would be challenging (but not impossible) to understand without the video. The 3D and 2D videos presented a pictorial representation of the ventricular system. As the audio lecture was being presented, the picture was dynamically changed to present the ventricle currently being discussed. Moreover, when a ventricle was being discussed, its name appeared on screen along with a line that connected the name with the relevant ventricle. The presentation of ventricles and names were synchronized identically for the 3D and 2D videos (i.e., names appeared with the same timing, placement, and format, as in the lecture only group).

Two-dimensional and 3D renderings were generated by an in-house software using the visualization toolkit (Schroeder, Martin, & Lorensen, 2006) and the ICBM 2009c Nonlinear Symmetric brain atlas data (Fonov, Evans, McKinstry, Almli, & Collins, 2009) acquired from the McConnell brain imaging center and preprocessed using ImageJ (Rasband & Image, 2016). Multiple custom color/opacity transfer functions and data clipping were used to generate stereoscopic imagery of the ventricular system from multiple viewpoints. Labeled stereoscopic viewpoints were acquired and montaged into a high resolution stereoscopic movie using the open-source video editing platform Shotcut (https://www.shotcut.org) with descriptive audio overdubbed. In addition, a version of the animation was created for the lecture only group, with only the audio description and text labels on the screen. The final movie was presented to the user using a passive 65 inch 4K LG3D TV (65UF8500) and LG passive eyewear either stereoscopically (3D), monoscopically (2D) or with only text labels and audio. The stereoscopic animation can be accessed at the following address: http://drosophila.biology.kent.edu/users/rclement/extras/ventricles.avi.

After filling out consent, we also administered a paper folding task as a single measure of spatial ability (e.g., Shepard & Feng, 1972), but its inclusion was exploratory and will not be discussed in any detail (until a brief mention in the General Discussion). For this task, a square piece of paper is depicted on a response sheet, and this virtual piece of paper is folded from 2 to 4 times; each fold is shown in sequence. After the final fold, a virtual whole is punched through the folded paper. Participants must mentally unfold the paper with the objective of being able to indicate where the holes appear in the unfolded paper; they are given five alternatives (unfolded squares with holes) to choose from (only one is correct). Participants had 3 min to complete 10 problems as accurately as possible.

For the cued-recall test, the 3D version of the video was shown again (a 2D version was also used in Experiment 2), and each ventricle was assigned a different number from 1 to 12. When a ventricle was presented, participants were instructed to do their best to write the name of the ventricle beside the corresponding number on a response sheet. For the matching test, the 3D version of the video was again presented, but the names of the ventricles were provided, and participants had to match the number of the ventricle (1 to 12 on the response sheet) with the correct ventricle name.

2.1.3 | Procedure

Participants were run either alone or in pairs, with any pair of participants being assigned to the same presentation group (because we had only one 3D television). After reading and filling out the consent form, the experiment was explained to them, and they were asked to do their best to learn the names of each ventricle. Participants sat approximately 7 feet from the television screen, and the lecture (with or without the video presentation) was presented 3 times (pilot data collected using the 3D version of the task indicated that 3 presentations produced mid-level test performance). After each presentation, a 15-s delay occurred while the experimenter (who sat in the same room) reset the lecture and video. After the final presentation, a 2-min delay occurred in which participants attempted to solve an engaging puzzle. Next, they predicted how many of the ventricles they would correctly label (out of 12) on the test. The tests were then administered, beginning with the cued-recall test (both tests are described in detail under Section 2.1.2). For both tests, participants could use as much time as needed to respond. Responses were hand scored and minor spelling errors were not counted against the total recall score.

2.2 | Results

In all experiments, for significant effects revealed by an analysis of variance (ANOVA), post hoc Tukey's tests were used to contrast means (as per recommendations from Jaccard & Guilamo-Ramos, 2002).

2.2.1 | Global judgment of learning

Means across participant's global judgments are presented in Table 1. A one-way ANOVA revealed a main effect of presentation format, $F(2, 87) = 4.98$, MSE = 0.04, $p = 0.01$, and partial $\eta^2 = 0.10$. Consistent with the multimedia-superiority hypothesis, global judgments were significantly lower for the group who listened to the lecture alone than for the 3D group, $p = 0.01$, Cohen's $d = 0.81$, although the post hoc test did not reach statistical significance for the 2D group, $p = 0.12$, $d = 0.53$. The difference in global judgments favored the 3D over the 2D group ($d = 0.24$), but this difference was not statistically significant, $p = 0.60$, $d = 0.24$.

2.2.2 | Test performance

The means across participant's average proportion of correct test performance are presented in Table 2, both for the cued recall test (which was administered first) and for the matching test. Given that the matching test was administered second and could have been influenced by the cued-recall test (i.e., the two tests were not independent), we limit the inferential analysis to the cued recall test (note, however, that the effects are similar on both measures). A one-way ANOVA on cued-recall performance revealed a main effect of presentation format, $F(2, 87) = 26.5$, MSE = 3.92, $p < 0.001$, partial $\eta^2 = 0.38$. Cued recall was not significantly greater for the 3D than 2D group, $p = 0.09$, $d = 0.52$. Performance was significantly lower for participants who listened to the lecture alone as compared with the 3D group, $p < 0.001$, $d = 1.82$, and the 2D group, $p < 0.001$, $d = 1.78$.

TABLE 1 Mean global judgment of learning

	3D presentation *M* (*SD*)	2D presentation *M* (*SD*)	Lecture only *M* (*SD*)
Experiment 1	0.48 (0.20)	0.43 (0.21)	0.33 (0.16)
Experiment 2			
3D test	0.60 (0.19)	0.51 (0.21)	—
2D test	0.64 (0.20)	0.58 (0.23)	—

Note. For Experiment 2, participants were not informed of the kind of test prior to making global predictions, so differences across test groups can be attributed to noise. 2D: two dimensional; 3D: three dimensional; *M*: mean; *SD*: standard deviation.

TABLE 2 Mean performance on the criterion test

	3D presentation *M* (*SD*)	2D presentation *M* (*SD*)	Lecture only *M* (*SD*)
Experiment 1			
Cued Recall	0.43 (0.22)	0.33 (0.12)	0.12 (0.12)
Matching	0.57 (0.27)	0.46 (0.19)	0.16 (0.12)
Experiment 2			
3D test	0.48 (0.23)	0.43 (0.19)	—
2D test	0.53 (0.25)	0.50 (0.24)	—

Note. For Experiment 2, only the matching test was used. 2D: two dimensional; 3D: three dimensional; *M*: mean; *SD*: standard deviation.

2.2.3 | Judgment accuracy

Analysis of judgment magnitude above was most relevant for addressing the main goal of the present research, which was to evaluate students' metamemory knowledge about 3D presentations. For interested readers, we also included analyses of the correspondence (or accuracy) of students' judgments to performance. We computed a difference score between each participant's predicted proportion of correct and actual proportion of correct on the cued-recall test. The signed difference score can reveal relatively good absolute accuracy on average even if deviations are large, because averaging positive (overconfident) and negative (underconfident) scores could result in a mean value that is close to 0. Thus, we also present absolute difference scores (or *unsigned* scores) to evaluate whether the absolute deviations between judgments and performance differ by groups. Both scores are presented in Table 3. A one-way ANOVA for the signed difference scores was significant, $F(2, 86) = 4.89$, MSE = 0.04, $p = 0.01$, partial $\eta^2 = 0.10$. Values were higher (indicating greater overconfidence) for the group who listened to the lecture alone than for the 3D group ($p = 0.002$, $d = 0.91$) and for the 2D group ($p = 0.04$,

TABLE 3 Measures of correspondence between judgments and memory performance

	3D presentation *M* (*SD*)	2D presentation *M* (*SD*)	Lecture only *M* (*SD*)
Experiment 1			
Signed	0.05 (0.22)	0.10* (0.20)	0.23* (0.18)
Unsigned	0.16* (0.15)	0.17* (0.15)	0.23* (0.17)
Correlation	0.49*	0.35	0.18
Experiment 2			
3D Test			
Signed	0.12* (0.18)	0.08 (0.25)	—
Unsigned	0.20* (0.17)	0.19* (0.15)	—
Correlation	0.67*	0.21	
2D Test			
Signed	0.10* (0.23)	0.08 (0.23)	—
Unsigned	0.21* (0.14)	0.17* (0.12)	—
Correlation	0.45*	0.46*	

Note. 2D: two dimensional; 3D: three dimensional; *M*:mean; *SD*: standard deviation. Signed: signed difference score between judgments and performance; Unsigned: absolute value of the signed difference scores. Correlation: Pearson *r* between judgments and recall.

*Value is significantly different from 0.

$d = 0.66$), whereas values were not significantly different for the two multimedia groups, $t(58) = 0.97$, $p = 0.77$. The group differences were diminished and not significant with the unsigned scores, $F(2, 86) = 1.49$, MSE = 0.04. Taken together, these outcomes suggest that participants who viewed only the lecture were generally overconfident, whereas participants in the multimedia groups demonstrated both over and underconfidence.

We also present correlations (r) across participants' judgments and recall performance in Table 3. These values reflect the degree to which individual differences in judgment magnitude correspond to individual differences in performance. All the correlations were positive, but only the correlation for the 3D group was significantly greater than 0. Note that the low value for the lecture only group may partly be an artifact of constrained variability in test performance (due to low overall performance) for that group. Also, given that these correlations were based on a small sample size (30 per group), any apparent difference among groups should be interpreted with caution.

2.3 | Discussion

Results from Experiment 1 support the conclusion that students believe that multimedia presentations improve learning. First, consistent with the multimedia-superiority hypothesis, students' judgments were higher when they viewed the 3D video with the lecture than when they listened only to the lecture without the supporting videos (the contrast involving the 2D video was in the same direction, albeit not statistically significant). Second, students' judgments tended to be higher for the 3D than 2D presentation, although this difference was not significant, so we estimated the size of this focal effect again.

3 | EXPERIMENT 2

In Experiment 2, we conducted a straightforward replication of Experiment 1 and also extended it as follows. First, the level of cued-recall performance was relatively low, so we decided to use only the matching test in Experiment 2. We also dropped the lecture only group, because replicating the medium-to-large effect sizes favoring multimedia over the lecture alone seemed trivial. Second, we manipulated the test format between participants: one group received the 3D test (as in Experiment 1) and the other received a 2D test. Including the 2D test allowed us to evaluate a transfer-appropriate-processing hypothesis (e.g., Blaxton, 1989; Morris, Bransford, & Franks, 1977), which would predict that test performance will be highest for those groups in which the encoding processes matched those required by the test. In the present case, performance on the 3D test is expected to be higher for the 3D than 2D presentation groups, whereas performance on the 2D test will be higher for the 2D than 3D presentation groups. Alternatively, 3D presentations may produce a more detailed representation of the target materials and subsequently support higher performance for both test formats. Finally, and most important, all groups predicted their performance on the upcoming test. We did not disclose the nature of the criterion test (i.e., whether it was a 3D or 2D test), because we did not want the test format to bias the students' judgments and potentially overshadow any impact of the 3D versus 2D presentations.

3.1 | Method

3.1.1 | Design, participants, materials, and procedure

Presentation format (2D vs 3D presentation) and test format (2D vs. 3D) were manipulated between participants. One hundred seventy-seven undergraduates from the same participant pool used in Experiment 1 (female percentage = 85; due to a communication error, ages were collected in only Experiment 1). Participants were randomly assigned to each of the three groups (ns = 40, 43, 42, and 42 for the 2D study-2D test, 2D study-3D test, 3D study-2D test, and 3D study-3D test groups, respectively). Given the small effect sizes reported in Experiment 1, we increased the sample size by 10 to potentially increase the statistical power to detect effects. Mean performance on the paper folding task did not significantly differ across groups: Ms were 4.8 ($SD = 2.6$), 5.2 (2.3), 5.2 (2.2), and 4.2 (2.3), for the 2D study-2D test, 2D study-3D test, 3D study-2D test, and 3D study-3D test groups, respectively, with all F values from a 2 × 2 ANOVA being less than 2.20. The first eight participants failed to provide a global prediction ($n = 2$ per group), but these participants were retained for estimating any format effects on criterion performance. The materials from Experiment 1 were used but also included a 2D formatted version of the test. The procedure was identical to the one used in Experiment 1, with two exceptions: Only the matching test was administered and some groups received a 2D version of the criterion test.

3.2 | Results

3.2.1 | Global judgment of learning

The mean across participant's global judgments are presented in Table 1. A 2 (presentation format: 3D vs. 2D) × 2 (test format: 3D vs. 2D) ANOVA revealed a main effect of presentation format, $F(1, 155) = 5.11$, MSE = 0.04, $p = 0.025$, partial $\eta^2 = 0.032$, with judgments being greater when students viewed the 3D than the 2D presentation (collapsed across test format, $d = 0.40$). The main effect of the test was less relevant because students were not told which test to expect when making the judgments, and it was not significant, $F(1, 155) = 3.49$, MSE = 0.04. The interaction was not significant, $F(1, 155) = 0.18$, MSE = 0.04.

3.2.2 | Test performance

The means across participant's proportion of correct test performance are presented in Table 2. A 2 × 2 ANOVA indicated that the main effect of presentation format was not significant, $F(1, 163) = 1.37$, MSE = 0.051, nor was the difference in test performance favoring the 2D test over the 3D test, $F(1, 163) = 3.42$, MSE = 0.051, $p = 0.07$. The interaction was not significant, $F(1, 163) = 0.04$, MSE = 0.051.

3.2.3 | Judgment accuracy

Measures of judgment accuracy are presented in Table 3. On average, participants tended to be overconfident. A 2 × 2 ANOVA was conducted separately for the signed and unsigned scores and revealed no significant main effects or interactions for either measure. The correlations between judgments and performance were also above 0 (although this correlation was not significantly different than 0 for the one group that received a 2D presentation and a 3D test). Given the small sample sizes per group, apparent differences among groups should be interpreted with caution.

3.3 | Discussion

In the first experiments, student's predictions tended to be higher for those who had viewed the 3D presentation rather than the 2D presentation. This difference in predictions was not significant in Experiment 1 but was significant in Experiment 2. To obtain an estimate of the effect size from these experiments, we used a continuously cumulating meta-analyses (CCMA) (Braver, Thoemmes, & Rosenthal, 2014) to combine outcomes from the replications. The pooled effect size (Cohen's *d*) was 0.36, $p < 0.01$, suggesting that the overall effect size is small to medium.[2] In Experiments 1 and 2, we manipulated the presentation format (3D vs. 2D) between participants because we were concerned that switching back and forth between a 3D and 2D format (as a within-participant manipulation) could be confusing and even increase the possibility of 3D sickness (e.g., dizziness from visual discomfort). Unfortunately, the use of a between-participant manipulation of presentation format may have limited the impact of students' beliefs about a 3D presentation on their global prediction. In particular, as noted in the introduction, judgments of learning have been linked to people's beliefs (for reviews, see Dunlosky, Mueller, & Tauber, 2015; Rhodes, 2016). For instance, many people believe that related word pairs (e.g., dog–cat) are easier to remember than are unrelated words pairs, and this belief is reflected in part by their judgments of learning (e.g., Mueller, Tauber, & Dunlosky, 2013; Undorf & Erdfelder, 2015). However, the majority of research investigating the effects of such factors (e.g., relatedness or any other variable) have manipulated them *within* each participant. When manipulated between participants, the influence of these factors on judgements of learning is reduced (e.g., Carroll, Nelson, & Kirwan, 1997; Dunlosky & Matvey, 2001; Shaddock & Carroll, 1997). Thus, the decision to manipulate presentation format between participants may have somewhat limited the impact of presentation format on participant's global judgments.

4 | SURVEY INVESTIGATION

Accordingly, to provide converging evidence for the conclusion that students believe that 3D (vs. 2D) presentations benefit learning more, each participant was first shown both presentations separately (e.g., the entire 3D presentation followed by the 2D presentation) and then was asked several questions. Most important, they were asked, “Do you think YOU learn more when watching 3D presentations of complex materials (e.g., Ventricular system) or with 2D presentations?” They answered on a 1 to 7 scale, with 1 indicating that they learn a lot more with 2D presentations, 4 indicating that they believe neither format is more effective, and 7 indicating that they learn a lot more with a 3D format. We also asked students other questions pertaining to the educational impact of 3D presentations (see Table 4 for the survey questions).

4.1 | Method

4.1.1 | Design, participants, materials, and procedure

The design included two groups that differed with respect to which video (3D or 2D) they viewed first. Forty-nine students (female percentage = 82) from the same pool used in the prior experiments participated to receive course credit and were randomly assigned to view the 3D presentation first ($n = 24$) or the 2D presentation first ($n = 25$). We had no a priori expectations for effect sizes, so we set our stopping rule relatively high ($N = 50$) but fell short by one participant (because the participant pool closed before we could complete the study). Mean performance on the paper folding task was 4.8 and 5.3 for the 3D first and 2D first groups, $t(47) = 0.83$, respectively. After viewing both videos, the participants filled out the survey (see Table 4). Given that participants viewed both presentations, we did not administer a criterion test.

4.2 | Results

Table 3 includes the mean survey response across participants, which are presented collapsed across all participants, and separately for those who viewed the 3D presentation first and for those who viewed the 2D presentation first. We also asked students how much experience they had watching 3D presentations (1 = *none* to 7 = *extensive*; *Median* = 4.0, *SD* = 1.6); reported experience with 3D presentations did not significantly correlate with the other responses so we do not discuss it further.

Several outcomes are notable in Table 4. First, all values indicate a preference of 3D over 2D presentations, whether it concerns learning more in general or viewing 3D presentations as a supplement to classroom lectures (*Median value* = 5.0 for both questions). Second, despite the preference toward 3D presentations, student's viewpoints in general were variable and not everyone endorsed the most extreme value (a rating of 7.0) for preferring 3D over 2D. For the question about preference (see Table 4, third question listed), approximately 30% of the ratings fell below 4.0, suggesting at least some students prefer 2D presentations. Understanding why these individual differences arise in preference will require further research. In summary, these results support the conclusion that the use of 3D materials for classroom learning will be viewed positively by many students.

5 | GENERAL DISCUSSION

The main goal of the project was to explore students' metamemory for multimedia presentations, with a focus on students' beliefs about the relative impact of viewing 3D versus 2D presentations. To summarize, consistent with the multimedia-superiority hypothesis, students judged that they learned more after multimedia presentations (involving a lecture and a corresponding video—either in 2D or 3D format)

---

[2] We also briefly reported the outcomes of an exploratory follow-up experiment involving a 2D and 3D group in the original submission. We were asked to exclude this experiment, but for full disclosure, the adjusted effect size from the CCMA including 2D versus 3D on judgment magnitude was .23 ($p = 0.05$) and consistent with a small effect size.

TABLE 4 Mean survey responses about 3D versus 2D presentation formats

Survey question	All participants M (SD)	2D first M (SD)	3D first M (SD)
Do you think you learn more when watching 3D presentation of complex materials or with 2D presentations?	4.8* (1.7)	5.1* (1.6)	4.5 (1.7)
Do you think people in general learn more when watching 3D presentation of complex materials or with 2D presentations?	5.0* (1.3)	5.0* (1.3)	5.0* (1.4)
If given a choice, would you prefer to learn from watching 3D presentations of complex materials or with 2D presentations?	4.8* (2.2)	5.3* (2.0)	4.3 (2.3)
To what degree do you think 3D presentations are a great format/tool for education?[a]	4.9* (1.6)	5.0* (1.7)	4.7* (1.5)
To what degree to you think 3D presentations are a great format/tool for entertainment?[a]	5.2* (1.8)	4.8* (1.8)	5.5* (1.7)
If you could view classroom materials that are visually complex, which format would you prefer using while studying alone?	4.4 (2.1)	5.0* (1.9)	3.8 (2.1)
If you could view classroom materials that are visually complex, which format would you prefer while viewing videos in the classroom?	5.0* (2.0)	5.4* (1.8)	4.6 (2.0)

Note. Students viewed both versions of the video (2D and 3D) immediately before completing the survey. 2D first: two-dimensional version of video presented before the 3D version; 3D first: three-dimensional version of video presented before the 2D version. 2D: two dimensional; 3D: three dimensional; *M*: mean; *SD*: standard deviation.

[a] Values for these questions were reversed scored so that higher values (1 to 7) indicated a positive response for using 3D presentations.

*$p < 0.05$, for a test against a value of 4.0 (indicating no preference).

than after listening to a lecture alone. In the present case, the lecture only group was presented with only the names of the ventricles for the same amount of time (and for the same duration) as those viewing the entire video, so the increased judgments (Table 1) for the multimedia presentations likely arise from presenting the ventricles themselves. That is, given the challenges of visualizing the ventricles when hearing the audio lecture, it perhaps is no surprise that students who also viewed the video presentation would perform better on the criterion test. Nevertheless, these outcomes critically extend prior research that investigated the impact of presenting static pictures with text material (e.g., Cardwell et al., 2017; Ikeda et al., 2013; Wiley et al., 2017) to a dynamic video in which the brain rotated to reveal each ventricle as it was discussed.

More important, based on the estimated effect size from a CCMA (see Section 3.3), students' made higher global judgments after viewing the 3D than 2D presentation. Moreover, most students reported that 3D presentations (vs. 2D ones) are more effective (Table 4). The impact of 3D versus 2D on judgments was also significant when the different kinds of video presentations (3D vs. 2D) had no impact on test performance (see Tables 1 and 2, Experiment 2), indicating that the differences in judgments are not based on direct access to how well the information had been encoded. Instead, we argue that these effects reflect people's beliefs about the potential efficacy of 3D presentations for improving learning. As noted in the introduction, such beliefs may arise from the different cues that are available when viewing 3D versus 2D presentations (e.g., the depth cues from 3D presentations), but another possibility is that students prior experience with 3D presentations have led them to believe that 3D presentations are superior. These and other alternatives need to be evaluated in future research, but regardless of how such beliefs arise, one implication is that many students may prefer 3D presentations even if they would *undermine* learning.

In the present context, however, test performance was consistently as good after the 3D than 2D presentation. Judgment accuracy (Table 3) was also at least as good for the 3D than 2D presentation. Perhaps most important, the 3D presentations did not increase student's overconfidence in their learning, which ultimately could undermine the effectiveness of their self-regulated learning (see Dunlosky & Rawson, 2012). Given these positive outcomes, teachers should consider using 3D presentations to support their classroom lectures on content that relies on depth information (see also, McIntire et al., 2014), with the following two *caveats*. First, a minority of participants appeared to have a preference for 2D formats over 3D formats (Section 4, Table 4), and understanding why these preferences arise will be important to make more fine-grained prescriptions of how to use 3D presentations in the classroom. For the survey study, we administered the paper folding task (Cronbach's $\alpha = 0.65$), and individual differences on this measure of spatial ability were not correlated with beliefs about learning for either one's own learning ($r = -0.07$) or for learning in general ($r = 0.07$). Thus, student's preferences for 3D multimedia do not appear to be related to their ability to internally manipulate shapes. Another possibility is that preference is related to students' overall comfort using 3D technology; according to the American Optometric Association, about 25% of people who view 3D videos experience negative symptoms, including eyestrain, headaches, or fatigue (from McIntire et al., 2014). Teachers should be aware that not all students may tolerate viewing 3D presentations and should have alternative plans when using this technology. The second *caveat* is that we suspect that the principles for effectively using 3D presentations will be similar to those for combining words with pictures. That is, multimedia presentations will be most effective when the 3D presentation highlights the essential to-be-learned concepts and is presented simultaneously with the lecture (for other principles of effective multimedia use, see Mayer, in press).

In summary, when college students attempted to learn the names of ventricles in the human brain, they (a) judged that 3D multimedia presentations are better than are lectures alone, (b) performed at least as well when the multimedia presentations included a 3D (vs. 2D) video, and (c) judged that 3D multimedia presentations were superior to 2D presentations. Individual differences arose in students' preference for viewing 3D (vs. 2D) multimedia presentations. Discovering why these differences occur and the degree to which they moderate the impact of 3D multimedia presentations on student learning are important goals for future research.

ACKNOWLEDGMENTS

For helping with data collection, we would like to thank Rachal Hall, Mason McLeodm, and Kayla Michael.

ORCID

John Dunlosky http://orcid.org/0000-0002-7367-7958

REFERENCES

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 657–668.

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, *9*, 333–342. https://doi.org/10.1177/1745691614529796

Brown, A., & Green, T. (2016). Virtual reality: Low-cost tools and resources for the classroom. *TechTrends*, *60*, 517–519. https://doi.org/10.1007/s11528-016-0102-z

Cardwell, B. A., Lindsay, S., Foerster, K., & Garry, M. (2017). Uninformative photos can increase people's perceived knowledge of complicated processes. *Journal of Applied Research in Memory and Cognition*, *6*, 244–252. https://doi.org/10.1016/j.jarmac.2017.05.002

Carroll, M., Nelson, T. O., & Kirwan, A. (1997). Tradeoff of semantic relatedness and degree of overlearning; differential effects on metamemory and long-term retention. *Acta Psychologica*, *95*, 239–253. https://doi.org/10.1016/S0001-6918(96)00040-6

Dongmei, C., Wilson, T. D., Rockhold, R. W., Lehman, M. N., & Lynch, J. C. (2017). Evaluatino of the effectiveness of 3D vascular stereoscopic models in anatomy instruction for first year medical students. *Anatomical Sciences Education*, *10*, 34–45.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1180–1191.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Beverly Hills, CA: SAGE.

Dunlosky, J., Mueller, M., & Tauber, S. K. (2015). The contribution of processing fluency (and beliefs) to people's judgments of learning. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger, III (Eds.), *Remembering: Attributions, processes, and control in human memory: Papers in honour of Larry L. Jacoby*. (pp.). 46–64). New York: Psychology Press.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*, 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Eitel, A. (2015). How repeated studying and testing affects multimedia learning: Evidence for adaptation to task demands. *Learning and Instruction*, *41*, 70–84.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Ferdig, R., Blank, J., Kratcoski, A., & Clements, R. (2015). Using stereoscopy to teach complex biological concepts. *Advances in Physiological Education*, *39*, 205–208. https://doi.org/10.1152/advan.00034.2014

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage, 47, Supplement 1.

Ikeda, K., Kitagami, S., Takahashi, T., Hattori, Y., & Ito, Y. (2013). Neuroscientific information bias in metacomprehension: The effect of brain images on metacomprehension judgment of neuroscience research. *Psychonomic Bulletin & Review*, *20*, 1357–1363. https://doi.org/10.3758/s13423-013-0457-5

Jaccard, J., & Guilamo-Ramos, V. (2002). Analysis of variance frameworks in clinical child and adolescent psychology: Issues and recommendations. *Journal of Clinical Child and Adolescent Psychology*, *31*, 130–146. https://doi.org/10.1207/S15374424JCCP3101_15

Jaeger, A. J., & Wiley, J. (2014). Do illustration help or harm metacomprehension accuracy? *Learning and Instruction*, *34*, 58–73. https://doi.org/10.1016/j.learninstruc.2014.08.002

Kaufmann, H., Schmalstieg, D., & Wagner, M. (2000). Construct3D: A virtual reality application for mathematics and geometry education. *Education and Information Technologies*, *5*, 263–276. https://doi.org/10.1023/A:1012049406877

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R. E. (in press). How multimedia can improve learning and instruction. To appear. In J. Dunlosky, & K. A. Rawson (Eds.), *The Cambridge Handbook on Cognition and Education*. New York: Cambridge University Press.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D displays and human performance: A comprehensive review. *Displays*, *35*, 18–26. https://doi.org/10.1016/j.displa.2013.10.004

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*, 378–384. https://doi.org/10.3758/s13423-012-0343-6

Perry, J. L., Cunningham, L. D., Gamage, J. K., & Kuehn, D. P. (2011). Do 3D stereoscopic computer animations improve student learning of surgical procedures? *International Journal of Instructional Media*, *38*, 369–378.

Petersson, H., Sinkvist, D., Wang, C., & Smedby, O. (2009). Web-based interactive 3D visualizations as a tool for improved anatomy learning. *Anatomical Sciences Education*, *2*, 61–68. https://doi.org/10.1002/ase.76

Rasband, W. S. & Image J. (1997-2016). U.S. National Institutes of Health, Bethesda, Maryland, USA. https://imagej.nih.gov/ij/

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky, & S. Tauber (Eds.), *The Oxford Handbook of Metamemory*. (pp. 65–80). NY, NY: Oxford University Press.

Schroeder, W., Martin, K., & Lorensen, B. (2006). *The visualization toolkit* (4th ed.)Kitware. ISBN 978-1-93093401901

Serra, M., & Dunlosky, J. (2010). Metacomprehension judgments reflect the belief that diagrams improve learning from text. *Memory*, *18*, 698–711. https://doi.org/10.1080/09658211.2010.506441

Shaddock, A., & Carroll, M. (1997). Influencues on metamemory judgements. *Australian Journal of Psychology*, *49*, 21–27. https://doi.org/10.1080/00049539708259846

Shepard, R. N., & Feng, C. (1972). A chronometric study of mental paper folding. *Cognitive Psychology*, *3*, 228–243. https://doi.org/10.1016/0010-0285(72)90005-9

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, *43*, 647–658. https://doi.org/10.3758/s13421-014-0479-x

Wang, H.-C., Chang, C.-Y., & Li, T.-Y. (2006). The comparative efficacy of 2D- versus 3D-based media design for influencing spatial visualization skills. *Computers in Human Behavior*, *23*, 1943–1957.

Wiley, J., Sarmento, D., Griffin, T. D., & Hinze, S. R. (2017). Biology textbook graphics and their impact on expectations of understanding. *Discourse Processes*, *54*, 463–478. https://doi.org/10.1080/0163853X.2017.1319655

How to cite this article: Dunlosky J, Dudley D, Spitznagel MB, Clements RJ. Student's metamemory knowledge about the impact of stereoscopic three-dimensional presentations of science content. *Appl Cognit Psychol*. 2019;33:225–233. https://doi.org/10.1002/acp.3469

APPENDIX A

The following is the transcript that comprised the lecture that corresponded to the video presentations.

The ventricular system is a system of linked spaces in which cerebrospinal fluid flows through the brain. This system is made up primarily of the two lateral ventricles, one inside each cerebral hemisphere, formerly called the first and second ventricles. Connected to the lateral ventricles is the third ventricle, and below the third ventricle is the fourth ventricle. Starting at the top are the largest ventricles, the two lateral ventricles. These ventricles mirror each other, and have a "c-shape" to them. The lateral ventricles have several extensions called horns that are named after the lobes into which they extend. The frontal horn of the lateral ventricle is the pointed aspect of this ventricle that extends toward the front part of the brain. Moving from the frontal horn toward the back of the brain is the body of the lateral ventricle. The body of the lateral ventricle extends toward the back of the brain, to the occipital horn of the lateral ventricle. Completing the c-shape, the ventricle extends forward toward the front of the brain as the temporal horn. The area in which the body, the occipital horn, and the temporal horn all meet is called the atrium of the lateral ventricle. Just below the frontal horn of the lateral ventricles is the interventricular foramen of Monro, a small channel that links the two lateral ventricles to the third ventricle. Below the third ventricle is the cerebral aqueduct of Sylvius, which connects the third ventricle to the fourth ventricle. Cerebrospinal fluid leaves the ventricular system via 3 openings. Two of these mirror each other, one extending to the right cerebral hemisphere, and the other to the left; these are called the foramina of Luschka. Finally, the largest of these openings is the foreman of Magendie, located below and extending from the fourth ventricle in the midline, into the central canal of the spinal cord.