Metacognition Learning (2018) 13:1–14
DOI 10.1007/s11409-017-9175-0

Instructor fluency leads to higher confidence in learning, but not better learning

Alexander R. Toftness[1] · **Shana K. Carpenter**[1] · **Jason Geller**[1] · **Sierra Lauber**[1] · **Madeline Johnson**[1] · **Patrick I. Armstrong**[1]

Received: 9 February 2017 / Accepted: 18 July 2017 / Published online: 28 July 2017

Abstract Students' judgements of their own learning often exceed their knowledge on a given topic. One source of this pervasive overconfidence is fluency, the perceived ease with which information is acquired. Though effects of fluency on metacognitive judgments have been explored by manipulating relatively simple stimuli such as font style, few studies have explored the effects of fluency on more complex forms of learning encountered in educational settings, such as learning from lectures. The present study manipulated the fluency of a 31-min video-recorded lecture, and measured its effects on both perceived and actual learning. In the fluent condition, the instructor used non-verbal gestures, voice dynamics, mobility about the space, and appropriate pauses. In the disfluent condition, the same instructor read directly from notes, hunched over a podium, rarely made eye contact, used few non-verbal gestures, spoke in monotone pitch, and took irregular and awkward pauses. Though participants rated the fluent instructor significantly higher than the disfluent instructor on measures of teaching effectiveness and estimated that they had learned more of the material, actual learning between the two groups did not differ as assessed by a memory test over the lecture contents given immediately (Experiment 1) or after a 1-day delay (Experiment 2). This counterintuitive result reveals an "illusion of learning" due to fluency in lecture-based learning, a very common form of instruction.

Keywords Instructor fluency · Metacognition · Learning · Overconfidence

Successful learning depends on the ability to accurately assess one's own knowledge. Understanding what is already known about a given topic is a necessary precursor to making effective study decisions (for a review, see Dunlosky and Tauber 2016). When students'

This material is based upon work supported by the James S. McDonnell Foundation twenty-first Century Science Initiative in Understanding Human Cognition, Collaborative Grant No. 220020483.

✉ Alexander R. Toftness
toftness@iastate.edu

[1] Department of Psychology, Iowa State University, W112 Lagomarcino Hall, Ames, IA 50011, USA

estimates of their own knowledge misalign with their actual knowledge, they are unlikely to regulate their study effectively and may either fail to devote adequate study time where it is needed, or devote too much study time where it is not needed.

The most common form of metacognitive misalignment is overconfidence. In a variety of studies, students' estimates of how much they know on a given topic often exceed what they actually know as measured by an objective memory test. This is true for participants in laboratory studies estimating their ability to remember simple stimuli such as word lists (e.g., Carpenter and Olson 2012; Dunlosky and Nelson 1994; Finn and Metcalfe 2007; Koriat and Bjork 2005), and for students in classrooms estimating their own performance on upcoming exams (e.g., Foster et al. 2017; Hacker et al. 2000; Hartwig and Dunlosky 2017; Miller and Geraci 2011). Understanding the factors contributing to overconfidence is important, as overconfidence could have the potential for detrimental consequences in educational settings. If students do not have a good grasp of what they know and do not know, for example, they may not develop the skills needed to effectively regulate their own learning.

One factor that can lead to overconfidence is fluency, or the ease with which information is perceived to be acquired during learning (for reviews see Alter and Oppenheimer 2009; Finn and Tauber 2015; Reber and Greifeneder 2017). Things that make something *appear* easy to learn sometimes increase confidence in learning without increasing actual learning. For example, students' judgments of their own learning—but not their actual learning—are often higher for visual words that are presented in a larger font size (Magreehan et al. 2016; Mueller et al. 2014; Rhodes and Castel 2008) or in a clearer font style (Yue et al. 2013), or for auditory words that are presented at an increased volume (Rhodes and Castel 2009). Students have also been known to overestimate their own learning of text material when it is presented in the context of colorful images, diagrams, and pictures (Carpenter and Olson 2012; Serra and Dunlosky 2010). The presence of animations in multimedia lessons has also been shown to increase students' confidence in their learning—particularly for students with lower proficiency on the topic—but does not affect their actual learning of the lesson content (Paik and Schraw 2012).

In academic settings, students encounter a variety of learning experiences that differ in fluency. Students may perceive the material in some courses as easier to understand than in other courses, some textbooks as better written and easier to read than others, and some instructors as more engaging, interesting, and easier to follow than others. A lively instructor who gives a smooth and well-polished lecture may give the impression that the lesson material is easy to take in and thus will be easy to remember, compared to an instructor who teaches the same lesson but in a way that is less engaging. Instructors vary in the degree of fluency they express while teaching, making lecture-based learning an important topic in educational research on metacognition. Currently, however, we know very little about the effects of lecture fluency on students' perceived and actual learning.

In the first known study to explore this under controlled laboratory conditions (Carpenter et al. 2013), participants watched one of two video-recorded lectures explaining the science behind calico cats. The two videos contained the same instructor, and the spoken content was scripted to ensure that the same material was presented. The only difference between the two videos was in the presentation style of the instructor. In the *fluent* video, the instructor stood upright, moved about the space, made eye contact with the camera, used appropriate gestures, and spoke seamlessly. In the *disfluent* video, the instructor hunched over a desk, stood in one place, read from notes, frequently broke eye contact, and spoke haltingly. Participants who watched the fluent video estimated that they learned significantly more of the content than participants who watched the disfluent video. However, a later test over the content presented in the videos revealed no significant difference in the amount of information learned between the two groups. Thus, participants in the fluent condition, but not in the disfluent condition, fell prey to an "illusion of learning," significantly overestimating how much they had learned.

Results of the Carpenter et al. (2013) study suggest that fluent instruction creates overconfidence in students, but has no benefit on actual learning. Given the intuitive assumption that instructors should try to be prepared, organized, and engaging, these results raise the troubling possibility that such behaviors may actually do a disservice to students by creating a false impression that the material has been well-learned when it really has not. A limitation of Carpenter et al.'s study, however, is that the videos were quite brief, lasting only about one minute. It is possible that fluent instruction does affect learning, but these effects are difficult to observe after one minute because such a brief interval does not allow participants enough time to become bored or disengaged.

Addressing this possibility, Carpenter et al. (2016) manipulated instructor fluency using a 22-min video that was created for an actual course. Akin to what students would experience in the classroom, the video contained visual displays and animated graphics to illustrate the concepts. Two versions of the same video were accompanied by the voice of an instructor explaining the scripted material in a way that was either fluent (deliberate, engaged, and confident) or disfluent (hesitant, disengaged, and uncertain). Similar to the results of Carpenter et al. (2013), participants' scores on a test following the videos did not differ according to which video they watched. In other words, fluent instruction again had no effect on learning. Importantly however, it also had no effect on participants' estimates of their own learning. Unlike Carpenter et al.'s (2013) results demonstrating an illusion of learning, the participants in Carpenter et al.'s (2016) study estimated their own learning to be similar following the fluent vs. disfluent lecture. However, follow-up analyses (Experiment 3) revealed that for participants who based their perceived learning on the instructor (compared to factors such as the material itself or their own learning abilities), participants who viewed the fluent instructor demonstrated the same illusion of learning as in Carpenter et al. (2013), whereas participants who viewed the disfluent instructor did not. Thus, it appears that the effects of instructor fluency depend, at least in part, on the degree to which the instructor is a salient factor influencing students' perceptions of their own learning.

In real lecture settings, the instructor is a salient part of the information delivery, being seen *and* heard by students. Realistic lectures also contain visual displays of the information being taught, increasing the chances that students have information other than the instructor (e.g., the organization of the visual graphics) on which to base their perceptions of learning. Previous studies on lecture fluency have included short videos with salient instructor presence (Carpenter et al. 2013), or longer videos containing visual graphics but no visual presence of the instructor (Carpenter et al. 2016). To date, previous studies have not explored the effects of instructor fluency under conditions in which the key components of a realistic lecture are all present—that is, when (a) the length of the videos is educationally-realistic, (b) the instructor is visually salient, and (c) the information being taught is displayed with the use of visual aids. Thus, the question remains open as to whether the fluency of instruction leads to differences in perceived and actual learning under realistic lecture conditions.

The current study was designed to explore the effects of instructor fluency using new materials with improved ecological validity. A 31-min video-recorded lecture on Information Theory was constructed, complete with visual aids and the presence of an instructor teaching the content. Two versions of the video were created, such that the instructor delivery style was either fluent or disfluent (see Fig. 1), but the visual graphics and spoken content were identical. As in previous research, participants' estimates of their own learning, as well as their actual learning, were assessed after watching one of the two videos.

To the extent that instructor salience influences students' estimates of their own learning (e.g., Carpenter et al. 2013, 2016, Experiment 3), participants in the current study might be inclined to believe they had learned more from the fluent instructor compared to the disfluent instructor. However, if participants base their perceived learning on something unrelated to the instructor (e.g., the lesson content as displayed in the visual aids), they may be less likely to show differences in perceived learning as a function of instructor fluency. Finally, as far as *actual* learning is concerned, if a salient, disfluent instructor encourages boredom and disengagement, then a 31-min video may provide sufficient time for participants to be affected by this, leading to decreased learning in the disfluent condition compared to the fluent condition.

Experiment 1

Participants

One-hundred and four undergraduate students from Iowa State University participated, and either received partial course credit or a payment of $10. Participants were tested individually on personal computers.

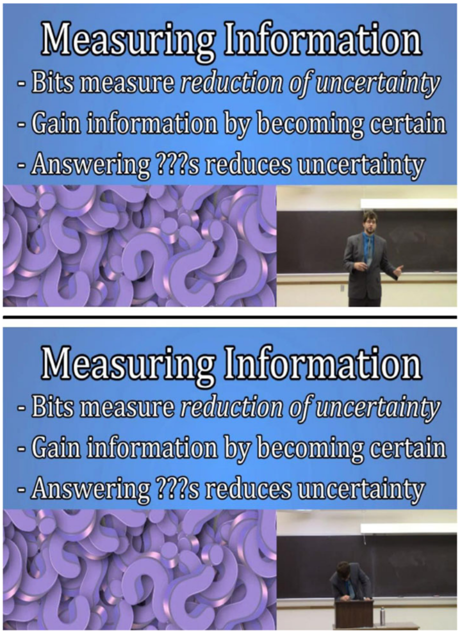

Fig. 1 Screenshots from the videos used in Experiments 1 and 2. The top panel depicts the fluent instructor, and the bottom panel depicts the disfluent instructor

Materials

Two versions of a video lecture on the topic of Information Theory were prepared.[1] Both versions contained the same PowerPoint slides and an inset of the instructor in the lower right-hand corner (see Fig. 1). The slides progressed at an identical rate in each version of the video, and the content was scripted so that the instructor spoke the same information across the two versions. Manipulated between the two video versions was the fluency of the delivery of the lecture. In the fluent condition (Fig. 1, top), the instructor used non-verbal gestures, voice dynamics, mobility about the space, and appropriate pauses. In the disfluent condition (Fig. 1, bottom), the instructor read directly from his notes, was hunched over a podium, rarely made eye contact, used few non-verbal gestures, spoke with a monotone pitch, and took irregular and awkward pauses. Each video lasted an identical length of 31 min.

A 30-item, four-alternative multiple-choice test was constructed from the video content. The questions were designed to assess fairly direct factual knowledge from the video, and each question had only one correct answer. To validate that the test was a fair assessment of knowledge from the video and that the answers were not easily guessable, the multiple-choice test was piloted on a unique sample of 15 participants (from the same participant pool as the experimental participants) who took the test *without* watching the video first. Average test performance for these 15 participants was 26.89% (SD = 6.7%), which was not significantly different from chance performance of 25%, $t(14) = 1.09$, $p = .30$.

Design and procedure

Participants gave informed consent before beginning the experiment. Participants were presented with instructions on the screen explaining that they were to view a lecture "about 30 minutes long" and to "watch the video carefully, and pay attention to the information that is being explained. Later, we will test your memory for this information." Participants were then instructed to use the headphones provided and begin viewing the video.

Participants were either shown the fluent version of the lecture (n = 52) or the disfluent version (n = 52). Immediately after the video ended, participants were asked to make a Judgement of Learning (JOL) pertaining to how well they believed they would perform on a multiple-choice test over the lecture material to be given after approximately one minute. Participants were instructed to enter a response from 0% to 100%. Immediately following the JOL response, participants were asked four questions pertaining to the perceived characteristics of the speaker in the video. On a scale of 1–5 (1 being "not at all" and 5 being "very"), they were asked to rate how organized the speaker was, how prepared the speaker was, how knowledgeable the speaker was, and the overall effectiveness of the speaker. Using the same 1–5 rating scale, participants then answered questions concerning how well they felt they had learned the material in the video, their level of interest in the information in the video, and their level of motivation to learn the information presented in the video.

Immediately after answering these questions, participants took the multiple-choice test over the material from the video. Questions were presented one at a time. The order of questions

[1] The lecture was created by the first author using a number of sources, and contained content relevant to the quantification, storage, and transmission of information. The topic was chosen because it was high in factual content and deemed unlikely to be highly familiar to undergraduate students taking a psychology course. References used in the construction of the lecture, the experimental script, and the multiple-choice test can be found at https://goo.gl/OxHBcs).

was identical for each participant and followed the order in which the information was presented in the video. The order of the four alternatives, A through D, was randomized for each question and each participant, ensuring that the correct answer was balanced across question alternative. An on-screen tracker displayed the number of questions participants had answered and the number remaining to be answered. Participants had unlimited time to respond to each question, with a 1000-ms delay in-between. Feedback was not provided after any of the questions.

Following the multiple-choice test, participants were asked if they had any prior knowledge of the information presented in the video. They were asked to respond with either "a) No, I did not have any detailed prior knowledge of this information," "b) I have heard of it before, but I did not know the details until today," "c) I may have learned this information before, but I did not remember the details," or "d) I learned this information before, and I remembered the details before coming to today's experiment." Following this, participants were debriefed, thanked, and dismissed from the experiment.

Results

Data from one participant in the fluent condition were excluded due to a computer malfunction that prevented the video from displaying properly. All analyses were conducted on the 103 remaining participants.

Instructor evaluation ratings The means and standard deviations of the instructor evaluation questions are given in Table 1. Participants rated the fluent instructor higher than the disfluent instructor on organization, $t(101) = 20.86$, $p < .001$, $d = 4.10$, knowledge, $t(101) = 13.88$, $p < .001$, $d = 2.75$, preparedness, $t(101) = 23.90$, $p < .001$, $d = 4.74$, and overall effectiveness, $t(101) = 13.50$, $p < .001$, $d = 2.65$. Participants also reported that they learned the material better from the fluent instructor ($M = 3.06$, $SD = .97$) compared to the disfluent instructor ($M = 1.90$, $SD = .89$), $t(101) = 6.30$, $p < .001$, $d = 1.24$. The fluent instructor, compared to the disfluent instructor, also garnered higher ratings of motivation (respectively, $M = 2.33$, $SD = .91$; $M = 1.63$, $SD = .95$), $t(101) = 3.81$, $p < .001$, $d = 0.75$, and interest (respectively, $M = 2.55$, $SD = .97$; $M = 1.85$, $SD = 1.13$), $t(101) = 3.40$, $p = .001$, $d = 0.67$.

Predicted versus actual performance Figure 2 shows mean test performance—both predicted (JOL) and actual—for the fluent vs. disfluent conditions. There was a significant difference in predicted performance between the two conditions, with participants in the fluent condition predicting significantly higher test scores ($M = 60\%$, $SD = 19\%$)

Table 1 Mean instructor evaluations (1–5) for the fluent and disfluent instructor

	Organized	Prepared	Knowledgeable	Effective
Experiment 1 (Immediate Test)				
Fluent ($n = 51$)	4.31 (.81)	4.78 (.46)	4.71 (.54)	3.75 (1.11)
Disfluent ($n = 52$)	1.48 (.54)	1.67 (.81)	2.31 (1.11)	1.40 (.57)
Experiment 2 (24-Hour Delayed Test)				
Fluent ($n = 39$)	4.36 (.81)	4.69 (.69)	4.69 (.57)	3.82 (1.02)
Disfluent ($n = 50$)	1.44 (.64)	1.44 (.76)	2.04 (.99)	1.48 (.81)

Standard deviations are given in parentheses. Data from Experiment 2 exclude one participant who completed the learning session and test session but did not provide answers to the instructor evaluation questions

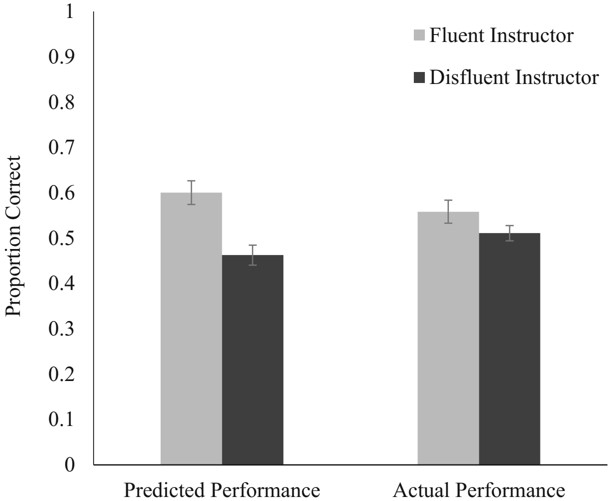

Fig. 2 Predicted vs. actual performance on the final test as a function of the fluent vs. disfluent instructor in Experiment 1. Error bars represent standard errors

than participants in the disfluent condition (M = 46%, SD = 18%), $t(101) = 3.79$, $p < .001$, $d = 0.75$. However, no significant differences occurred in actual test performance between the fluent (M = 56%, SD = 16%) and disfluent (M = 51%, SD = 12%) conditions, $t(101) = 1.70$, $p = .091$.[2]

A 2 × 2 (Performance: Predicted vs. Actual x Condition: Fluent vs. Disfluent) Mixed Analysis of Variance (ANOVA) confirmed that this interaction was significant, $F(1, 101) = 5.17$, $p = .025$, $\eta^2 = .049$. The main effect of condition was also significant, $F(1, 101) = 13.20$, $p < .001$, $\eta^2 = .12$, in that the fluent condition was associated with overall higher performance (collapsed across predicted and actual) than the disfluent condition. The main effect of performance (predicted vs. actual) was not significant, $F = .02$, indicating that participants were not generally overconfident in their learning.

Prior knowledge Most participants (62%) reported having no prior knowledge of the information in the video. Some reported having heard of it before (15%), and some reported that they may have learned it before but did not remember the details prior to the experiment (23%). No participants reported having detailed prior knowledge of the material at the time of the experiment. To control for any effects of prior knowledge, the same 2 × 2 ANOVA from above was run including only data from participants who reported having no prior knowledge of the material whatsoever. The analysis based on these participants (n = 64) yielded the same interaction, $F(1, 62) = 8.97$, $p = .004$, $\eta^2 = .13$, whereby JOLs were higher for the fluent vs. disfluent condition (60% vs. 40%, respectively), $t(62) = 4.24$, $p < .001$, $d = 1.08$, but test scores did not differ (53% vs. 48%, respectively), $t(62) = 1.29$, $p = .202$. Thus, even when prior knowledge was controlled, the fluent instructor led to greater confidence in learning, but did not increase actual learning.

[2] Test performance in both conditions was significantly higher than the performance of the 15 participants who took the test without watching the video first (see Materials section of Experiment 1). This was true for both Experiment 1 and Experiment 2, ts > 4.65, ps < .001, indicating that significant learning occurred in both the fluent and disfluent conditions.

Discussion

The fluent instructor, rated higher than the disfluent instructor on traditional assessments of teaching effectiveness, produced higher confidence in participants' learning, but did not produce higher actual learning. The tendency for the fluent instructor to produce an "illusion of learning" is consistent with the findings from previous studies on instructor fluency using short videos (Carpenter et al. 2013) and videos that manipulated the vocal cues, but not the visual appearance, of the instructor (Carpenter et al. 2016). Using materials that are more educationally realistic—over 30 min in length, involving visual aids and salient instructor cues—Experiment 1 confirms that fluent instructors may increase confidence in learning, but do not appear to increase actual learning. Taken together, past and present results provide five experiments that have failed to demonstrate any significant effects of instructor fluency on objective measures of learning.

One exception to this was Carpenter et al.'s (2016) Experiment 3, which revealed a small but significant advantage of the fluent instructor over the disfluent instructor on test scores when the test was given after a 24-h delay. Given that the other experiments on this topic have involved fairly immediate tests, it is unknown whether consistent effects of instructor fluency show up on (more educationally-relevant) delayed tests of memory. If disfluent instruction leads to less durable learning, or faster forgetting, than fluent instruction, these effects would be more likely to appear on delayed assessments. Thus, Experiment 2 was a replication of Experiment 1, this time with the multiple-choice test delayed by 24 h.

Experiment 2

Participants

One-hundred and nine new undergraduate students from Iowa State University participated, and received either partial course credit or a payment of $10. Participants were tested individually on personal computers.

Materials, design and procedure

The materials, design, and procedure were identical to Experiment 1, except that the multiple-choice test was administered one day after the learning session. As in Experiment 1, participants watched one of the two videos, then completed the JOL question (asking them to estimate their performance on a multiple-choice test to be given the next day), then completed the same seven questions assessing instructor effectiveness and self-reported level of the information learned, interest, and motivation. Upon returning for the second session, participants were given the same multiple-choice test from Experiment 1. Immediately afterward, they were asked to indicate their prior knowledge of the material in the video, and then were asked whether they had looked up, rehearsed, shared, or otherwise re-exposed themselves to any of the information from the video since completing the learning session the previous day.

Results

Data were excluded from 12 participants who failed to return for the second session (two participants in the disfluent condition, and ten in the fluent condition), three participants who experienced computer malfunctions that prevented the video from displaying properly (one in the disfluent condition and two in the fluent condition), one participant in the disfluent condition who fell asleep during the learning session, one participant in the fluent condition who reported having high prior knowledge of the information in the video (along with a nearly perfect test score) and knowing the details prior to participating in the study, and from two participants in the fluent condition who reported looking up or discussing the information in the video in-between the learning session and the test session. Data from the 90 remaining participants were analyzed.

Instructor evaluation ratings The means and standard deviations of the instructor evaluation questions are given in Table 1. As in Experiment 1, participants rated the fluent instructor higher than the disfluent instructor on organization, $t(87) = 18.94$, $p < .001$, $d = 3.99$, knowledge, $t(87) = 14.92$, $p < .001$, $d = 3.28$, preparedness, $t(87) = 20.80$, $p < .001$, $d = 4.47$, and overall effectiveness, $t(87) = 12.02$, $p < .001$, $d = 2.53$. Participants also reported that they learned the material better from the fluent instructor ($M = 3.50$, $SD = .98$) compared to the disfluent instructor ($M = 2.06$, $SD = .91$), $t(86) = 7.10$, $p < .001$, $d = 1.52$. The fluent instructor, compared to the disfluent instructor, also garnered higher ratings of motivation (respectively, $M = 2.62$, $SD = 1.02$; $M = 1.58$, $SD = .86$), $t(87) = 5.21$, $p < .001$, $d = 1.10$, and interest (respectively, $M = 2.85$, $SD = 1.23$; $M = 1.86$, $SD = 1.20$), $t(87) = 3.82$, $p < .001$, $d = 0.81$.

Predicted versus actual performance Figure 3 shows mean test performance—both predicted (JOL) and actual—for the fluent vs. disfluent conditions. As in Experiment 1, there was a significant difference in predicted performance between the two conditions, with participants in the fluent condition predicting significantly higher test scores ($M = 66\%$, $SD = 19\%$) than participants in the disfluent condition ($M = 50\%$, $SD = 25\%$), $t(88) = 3.32$, $p = .001$, $d = 0.71$. However, also as in Experiment 1, no significant differences occurred in actual test performance between the fluent ($M = 48\%$, $SD = 16\%$) and disfluent ($M = 45\%$, $SD = 14\%$) conditions, $t(88) = 1.18$, $p = .241$.[2]

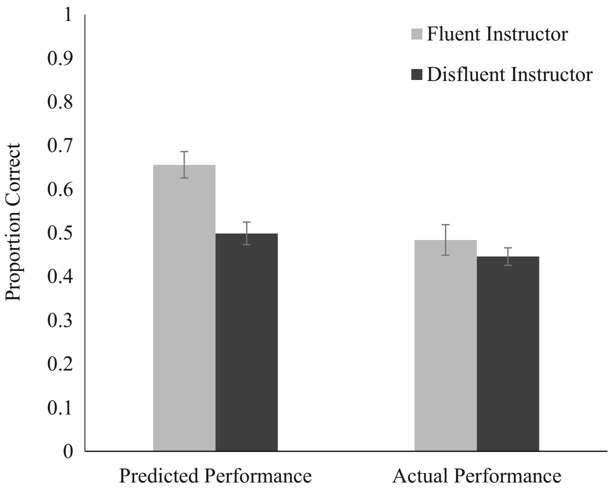

Fig. 3 Predicted vs. actual performance on the final test as a function of the fluent vs. disfluent instructor in Experiment 2. Error bars represent standard errors

A 2 × 2 (Performance: Predicted vs. Actual x Condition: Fluent vs. Disfluent) Mixed Analysis of Variance (ANOVA) confirmed that this interaction was significant, $F(1, 88) = 8.32$, $p = .005$, $\eta^2 = .086$. The main effect of condition was also significant, $F(1, 88) = 7.84$, $p = .006$, $\eta^2 = .08$, in that the fluent condition was associated with overall higher performance (collapsed across predicted and actual) than the disfluent condition. The main effect of performance (predicted vs. actual) was significant, $F(1, 88) = 29.58$, $p < .001$, $\eta^2 = .25$, indicating that participants were generally overconfident in their learning.

Prior knowledge Most participants (62%) reported having no prior knowledge of the information in the video. Some reported having heard of it before (27%), and some reported that they may have learned it before but did not remember the details prior to the experiment (11%). As in Experiment 1, the same 2 × 2 ANOVA from above was run including only data from participants who reported having no prior knowledge of the material. This analysis ($n = 56$) revealed the same interaction, $F(1, 54) = 9.23$, $p = .004$, $\eta^2 = .15$, whereby JOLs were higher for the fluent vs. disfluent condition (65% vs. 44%, respectively), $t(54) = 3.38$, $p = .001$, $d = 0.93$, but test scores did not differ (49% vs. 44%, respectively), $t(54) = 1.20$, $p = .234$. Thus, as in Experiment 1, after controlling for prior knowledge, the fluent instructor led to greater confidence in learning without increasing actual learning.

Discussion

Experiment 2 confirmed that actual test performance was not significantly different between the fluent vs. disfluent conditions. This was true despite the instructor in the fluent condition being associated with higher ratings of teaching effectiveness and higher judgments of learning. Thus, the illusion of learning, driven by instructor fluency, appears to exist whether the test is given fairly immediately after learning (Experiment 1), or after a 24-h delay (Experiment 2).

General discussion

The current study provides new data on the effects of instructor fluency on student learning. Participants who viewed the fluent instructor, compared to those who viewed the disfluent instructor, estimated that they had learned more of the information from a video-recorded lecture, but failed to show an advantage in actual learning as measured by an objective multiple-choice test over the video's contents.

This finding is consistent with previous research using shorter lecture videos of only one minute (Carpenter et al. 2013). This "illusion of learning" indicates that students can base their perceptions of their own learning on the fluency or ease with which an instructor presents information, and this appears to be true whether the lecture video is very short on the order of one minute, or a more educationally-realistic duration of 31 min.

It was hypothesized that a 31-min video, delivered by a disfluent instructor, might encourage boredom or disengagement among participants, leading to a learning decrement relative to the same content delivered by a fluent instructor. However, we found no significant differences in actual learning, as measured by an objective multiple-choice test, between participants who viewed the fluent instructor vs. the disfluent instructor. This finding is consistent with previous studies on instructor fluency (Carpenter et al. 2013, 2016), and adds to a small but growing literature showing—somewhat counterintuitively—that fluent instruction does not appear to affect the amount of information that participants learn.

The current study rules out potential floor effects as one explanation for this finding. Previous studies on instructor fluency have sometimes observed fairly low performance on the final criterial test (only about 25% accuracy on the open-ended test in the study by Carpenter et al. 2013), raising the question of whether the difficulty of learning and remembering the material could have obscured differences in test performance between the fluent and disfluent conditions. In the current study, we collected base-rate data on the probability of guessing the correct answers to the multiple-choice test. We found that the guess rate was no different from chance performance, and was significantly lower than the test performance of participants in both the fluent and disfluent conditions after watching the videos. This was true under conditions where the test occurred immediately (Experiment 1) or after one day (Experiment 2). Thus, we can verify that significant learning occurred by watching the videos in both the fluent and disfluent conditions, and that the amount of learning that occurred did not differ by condition.

The lecture videos used in the current study were more ecologically valid than those used in previous research (Carpenter et al. 2013, 2016). This adds to the reliability of these findings under conditions where the duration of the lecture was educationally-relevant, the instructor was salient, and the instructor used visual graphics and animations to explain the content as is commonplace in everyday modern lecture settings. Also consistent with real classroom situations, we found that participants sometimes possessed a range of self-reported prior knowledge of the material presented in the videos. Such a situation is akin to a classroom setting in which students often arrive with some pre-existing ideas about the topic, or perhaps have already taken a class related to the topic. Importantly, we found that the illusion of learning due to fluency—the key result observed in both experiments—occurred independently of participants' levels of prior knowledge.

Thus, fluency of instruction does not appear to influence the processing of information in a way that affects learning in the short-term (Experiment 1) or long-term (Experiment 2), at least over the intervals of time measured here. The consistency of this finding over different retention intervals is of theoretical interest, as it suggests that instructor fluency affects neither immediate learning, nor the durability of learning over time. Fluency of instruction does consistently affect students' *perceptions* of their learning, however. Whether the test was administered immediately or after a delay, students believed they had learned more from the fluent instructor compared to the disfluent instructor. The "illusion of learning" based on fluency therefore appears to exist under conditions of both immediate and delayed assessment.

What might account for this illusion of learning? With regards to the mechanism(s) underlying students' judgments of learning (JOLs), recent research has distinguished between subjective experiences based on ease of processing, vs. pre-existing beliefs about learning (e.g., Mueller et al. 2013). When perceptual stimuli are manipulated such that one condition is noticeably different or stands out in some way—e.g., the font size is larger (Magreehan et al. 2016), the volume is louder (Rhodes and Castel 2009), or the lecture delivery is smoother (Carpenter et al. 2013), participants might perceive that they have learned it better because of either some subjective experience associated with processing the stimuli (i.e., they can read it faster), or because of a pre-existing belief that things are learned better when they are bigger, louder, or smoother.

Evidence is accumulating that JOLs more often reflect belief-based decisions than experience-based decisions (e.g., Mueller et al. 2014, 2016, 2013). Research on instructor fluency has not yet collected overt responses (e.g., speed of processing) to measure the subjective experience of ease of processing while learning from fluent vs. disfluent lectures. To the extent that such overt measures do not differ as a function of lecture fluency, it is likely that the higher JOLs associated with fluent instructors compared to disfluent instructors reflect students' beliefs that fluent instructors are better teachers and thus lead to better learning. Future research is encouraged that can directly measure the processing that students engage in while learning from lectures (e.g., via response probes or note-taking), the role that the instructor plays, and factors that may mitigate the illusion of learning driven by content-independent factors such as instructor delivery style.

Participants' perceptions of the instructor clearly differed as a function of the fluency of delivery. Participants rated the fluent instructor significantly higher than the disfluent instructor on traditional measures of teaching effectiveness, including organization, preparedness, knowledge, and overall effectiveness. This is consistent with previous research on instructor fluency using the same paradigm as in the current study (Carpenter et al. 2013, 2016). This finding also parallels results from earlier studies showing that an instructor's degree of expressiveness (e.g., use of humor and personal anecdotes) coincides positively with the effectiveness ratings they receive, but does not always coincide with the amount of information students learn (e.g., Perry et al. 1979; Ware and Williams 1975; Williams and Ware 1976, 1977). These findings suggest that student evaluations of instructors may not be the most reliable indicators of learning.

Indeed, research has shown that instructor evaluations can be influenced by a number of factors irrelevant to the material being learned, such as an instructor's expressiveness, age, gender, personality, and attractiveness (Abrami et al. 1982; Neath 1996; Spooren et al. 2013; Ware and Williams 1975; Williams and Ceci 1997). Along these lines, one recent laboratory study found that an instructor who spoke with accented English, compared to a native English-speaking instructor who presented identical content, was rated lower on measures of instructional quality, even though participants learned the material equally well from the two instructors (Sanchez and Khan 2016). In their recent review of college instructor evaluations, Kornell and Hausman (2016) reported preliminary data that student evaluations of instructor effectiveness might even be negatively correlated with long-term learning as measured in follow-up courses.

The results of the current study are limited by the fact that participants' learning was assessed in the laboratory after viewing only one video-recorded lecture. It could be that in actual courses, fluent instruction is associated with better learning, due perhaps to extended time spent with the instructor (including in-class and out-of-class meetings), and the possibility that fluent instructors encourage behaviors among students that might indirectly benefit learning, such as increased attendance and more time spent studying outside of class.

In the only known study to address this, the relationship between instructor fluency and learning in actual courses was recently explored by Serra and Magreehan (2016). On an end-of-semester survey administered to multiple sections of introductory psychology, students were asked to rate their instructors on several fluency-related attributes (e.g., the degree to which the instructors spoke clearly, maintained students' attention, used visual aids, etc.), and to estimate how much of the material they felt they had learned from the course (0–100%). Although there was a small correlation between students' ratings of instructor fluency and their final grades in the course ($r = .13$), after controlling for final grades, fluency ratings correlated significantly with students' estimates of how much they felt they had learned from the course ($r = .38$, $p < .001$), and with their overall evaluation of both the instructor ($r = .74$, $p < .001$) and the course ($r = .38$, $p < .001$). These important classroom data indicate that under the most ecologically valid conditions, students' perceptions of the fluency of their instructors relate significantly to their estimates of how much they feel they have learned and their overall impressions of the course.

Future research would benefit from further explorations of the role of instructor fluency in realistic classroom situations, and the mechanisms behind fluency-based metacognitive decisions in education. Given that fluency-based decisions occur in real educational settings (Serra and Magreehan 2016), future research is also encouraged that explores the effects of fluency on outcomes of consequence to education. Though instructor fluency did not significantly affect test scores in the current study, it is possible that instructor fluency affects other outcomes not measured here, such as students' interests to learn more about the topic, decisions to take more courses in the area, approaches to studying, or choices of academic major. Positive experiences with instructors could have downstream benefits for learners—both subjective and objective—that are important to their educational experiences and worth exploring in future research.

Compliance with ethical standards

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

References

Abrami, P. C., Leventhal, L., & Perry, R. P. (1982). Educational seduction. *Review of Educational Research, 52*, 446–464.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235.

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 92–101.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review, 20*(6), 1350–1356.

Carpenter, S. K., Mickes, L., Rahman, S., & Fernandez, C. (2016). The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. *Journal of Experimental Psychology: Applied, 22*(2), 161–172.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when JOLs occur? *Journal of Memory and Language, 33*, 545–565.

Dunlosky, J., & Tauber, S. K. (Eds.). (2016). *The Oxford handbook of Metamemory*. New York: Oxford University Press.

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 238–244.

Finn, B., & Tauber, S. K. (2015). When confidence is not a signal of knowing: How students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educational Psychology Review, 27*(4), 567–586.

Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition and Learning*, *12*(1), 1–19.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160–170.

Hartwig, M. K., & Dunlosky, J. (2017). Category learning judgments in the classroom: Can students judge how well they know course topics? *Contemporary Educational Psychology, 49*, 80–90.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 187–194.

Kornell, N., & Hausman, H. (2016). Do the best teachers get the best ratings? *Frontiers in Psychology, 7*, 570.

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning, 11*, 35–56.

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 502–506.
Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness judgments of learning. *Psychonomic Bulletin & Review, 20*, 378–384.
Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70*, 1–12.
Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *Quarterly Journal of Experimental Psychology, 69*, 781–799.
Neath, I. (1996). How to improve your teaching evaluations without improving your teaching. *Psychological Reports, 78*, 1363–1372.
Paik, E. S., & Schraw, G. (2012). Learning with animation and illusions of understanding. *Journal of Educational Psychology, 105*(2), 278–289.
Perry, R. P., Abrami, P. C., & Leventhal, L. (1979). Educational seduction: The effect of instructor expressiveness and lecture content on student ratings and achievement. *Journal of Educational Psychology, 71*, 107–116.
Reber, R., & Greifeneder, R. (2017). Processing fluency in education: How metacognitive feelings shape learning, belief formation, and affect. *Educational Psychologist, 52*(2), 84-103.
Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615–625.
Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16*(3), 550–554.
Sanchez, C. A., & Khan, S. (2016). Instructor accents in online education and their effect on learning and attitudes. *Journal of Computer Assisted Learning, 32*, 494–502.
Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgements reflect the belief that diagrams improve learning from text. *Memory, 18*(7), 698–711.
Serra, M. J., & Magreehan, D. A. (2016). Instructor fluency correlates with students' ratings of their learning and their instructor in an actual course. *Creative Education, 7*(8), 1154–1165.
Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 10*, 1–45.
Ware Jr., J. E., & Williams, R. G. (1975). The Dr. Fox effect: a study of lecturer effectiveness and ratings of instruction. *Journal of Medical Education, 50*, 149–156.
Williams, W. M., & Ceci, S. J. (1997). "How'm I doing?" problems with student ratings of instructors and courses. *Change: The Magazine of Higher Learning, 29*(5), 12–23.
Williams, R. G., & Ware, J. E., Jr. (1976). Validity of students' ratings of instruction under different incentive conditions: A further study of the Dr. Fox effect. Journal of Educational Psychology*, 68,* 48–56.
Williams, R. G., & Ware Jr., J. E. (1977). An extended visit with Dr. fox: Validity of student satisfaction with instruction ratings after repeated exposures to a lecture. *American Educational Research Journal, 14*, 449–457.
Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition, 41*(2), 229–241.