Received: 5 December 2022 | Revised: 5 April 2023 | Accepted: 10 June 2023

DOI: 10.1002/acp.4101

RESEARCH ARTICLE

WILEY

The forward testing effect after a 1-day delay across dissimilar video lessons

Jocelyn Parong | **C. Shawn Green**

Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin, USA

Correspondence
Jocelyn Parong, Department of Psychology, University of Wisconsin–Madison, Madison, WI, USA.
Email: parong.jocelyn@gmail.com

Abstract

Being tested on previously learned material has been shown to enhance the learning of subsequently encountered material (i.e., the forward testing effect). Some cognitive theories predict that the magnitude of the forward testing effect is not dependent on the content in the previously learned and to-be-learned material (content-general theories), while other theories predict that it does depends on the similarity (content-specific theories). To adjudicate between these broad theories, participants viewed one of four lessons that were similar or dissimilar on two dimensions, academic domain and presentation structure, relative to a second lesson. After the first lesson, participants were either tested or restudied the material. Then they viewed the second lesson and were tested on the material. Results showed an overall forward testing effect across all four lessons, and the magnitude of the effect was not significantly different between the lessons, providing support for content-general theories of the forward testing effect.

KEYWORDS

forward testing, interim testing, retrieval practice, testing effect

1 | INTRODUCTION

The word "testing" perhaps most commonly evokes thoughts regarding the assessment of learning that has already taken place (e.g., a standardized test, a mid-term, a final exam, etc.). Yet there is a now a sizeable body of research demonstrating that testing can be used to enhance the learning process itself. Indeed, there are many published reports in which participants are first asked to study a set of learning material (e.g., a list of words or some academic content). Half of the participants are then asked to restudy the material, while the other half are given a test on the material. Finally, after some delay, both groups are then given a final test on the material. The consistent finding—dubbed the "*testing effect*"—is that, on the final test, those participants who were given a test at the intermediate stage outperform the participants who were asked to restudy (Roediger III & Karpicke, 2006).

More recently, an extension of this basic paradigm has revealed a second way that testing can enhance learning. The basic procedure begins identically as above, with all participants first studying some material, followed by half of the participants being asked to restudy the material and the other half being tested on the material. Now though, rather than all participants then receiving a final test on the material, the participants are instead given some brand-new second material to learn before finally receiving a final test on this new second material. Here the typical finding has been that those participants who received an intermediate test on the first material outperform the participants who restudied the first material on the final test of the new second material. This outcome has been dubbed the "forward testing effect" (also referred to as test-enhanced learning, test-potentiated new learning, or the interim testing effect). In essence, being tested on previously learned material serves to enhance the learning of subsequently encountered material

(Chan et al., 2018; Pastötter & Bäuml, 2014; Yang et al., 2018; Yang, Luo, et al., 2021).

Evidence for the forward testing effect has been found using several types of learning material. These include lists of single words (e.g., Szpunar et al., 2008; Weinstein et al., 2014; Yang et al., 2017), word pair associates (e.g., Weinstein et al., 2011), text passages (e.g., Wissman et al., 2011), video lectures (e.g., Jing et al., 2016; Szpunar et al., 2013), category learning (e.g., Lee & Ahn, 2018; Yang & Shanks, 2018), spatial information (Bufe & Aslan, 2018), and motor learning (Tempel & Frings, 2019). One meta-analysis found an overall moderate forward testing effect ($g = 0.44$), and for lessons consisting of videos or prose passages, the effect was even larger ($g = 0.77$; Chan et al., 2018). A more recent meta-analysis of multi-list forward testing effect paradigms, which corrected for publication bias, found a larger, more robust average effect ($g = 0.75$; Boustani & Shanks, 2022).

Given the meta-analytic work suggesting that the forward testing phenomenon is a reasonably robust empirical finding, several theories have been proposed to explain the mechanism through which the forward testing effect occurs. The proposed explanations can be generally grouped into two broad flavors: (1) content-general theories, which assume the mechanisms at play are independent of the specific content in either the initially learned and tested or the subsequently learned lessons (or the interaction between the content in the two lessons) and (2) content-specific theories, which propose that the underlying mechanisms of the forward testing effect include specific changes in attention and/or strategy that are dependent on the similarity between the lessons on at least some dimensions.

Many content-general theories of the forward testing effect, at their core, propose that there are changes in global attentional processes that are spurred by testing of previously studied material. For example, interim tests may serve as a marker to students that they had not previously maintained a sustained level of attention (i.e., lapses in attention, often manifested as mind-wandering, are known to increase throughout the duration of a lesson; Risko et al., 2013). This then spurs deeper attention in a subsequent learning experience (Szpunar et al., 2013). Alternatively, tests may provide a sense of cognitive closure, reducing the desire to hold previously learned material in working memory and freeing up attentional resources to focus on the new learning (Roets et al., 2006). Similarly, the reset-of-encoding hypothesis of the forward testing effect posits that testing of previously studied information resets the encoding process and makes the encoding of new information more effective (Pastötter et al., 2011, 2018). Metacognitive theories of the forward testing effect meanwhile argue that after taking a test, learners may be more cognizant of the need to monitor their understanding during the new learning block in order to perform well on a subsequent test. In a similar vein, testing may act as a calibrator of confidence during learning; without overt retrieval, learners are often overconfident about how much they will subsequently remember from a lesson (Dunlosky & Rawson, 2012). Finally, testing may enhance subsequent learning by increasing motivation and effort during the succeeding learning block. This increase in effort may be due to awareness of retrieval failures during testing. After an initial test, participants were more likely to put more effort into items that were perceived as more difficult (Soderstrom & Bjork, 2014). Moreover, exerting more effort during the second learning phase may be due to test expectancy as the presence of a test increases the expectation of a test after subsequent blocks of learning (Chan et al., 2018; Yang et al., 2018). While none of these content general theories are mutually exclusive (indeed, some or all of these mechanisms could conceivably work in concert), these content-general theories implicitly posit that the forward testing effect should be comparable regardless of whether the new material is similar or dissimilar to the previously tested material.

Content-general theories have found support in results, for instance, showing that a forward testing effect appears not only in cases where the initial and subsequent learning material are quite similar in nature, but also across learning material that differs along some dimension. For example, Yang et al. (2019) contrasted the forward testing effect between a lesson that presented factual statements about painters (declarative knowledge) and a lesson of categorizing paintings by artists (category learning of perceptual features). The authors found that testing after the declarative knowledge lesson increased learning of the category learning material compared to restudying the declarative knowledge lesson. In another study, Yue et al. (2015) found a forward testing effect between two lessons in different academic domains—astronomy and meteorology; participants who were tested after viewing a lesson on the life cycle of a star performed better on a subsequent test after a lesson on the formation of lightning compared to participants who re-studied the first lesson.

However, it could be argued that the lessons used in the two studies above still share at least one dimension of commonality which could, in turn, be the true driving force underlying the forward testing effect. Indeed, although learning declarative facts about artists may differ from learning visual features of paintings to later categorize them in how the lesson is structured and/or how the information is organized and stored, they are both still within the art domain (Yang et al., 2019). Similarly, though an astronomy lesson and meteorology lesson span different academic domains, the lessons used in this study both involve the presentation of information in a procedural manner (e.g., steps of the star life cycle and steps of how lightning forms; Yue et al., 2015).

As such, these results have not fully ruled out more content-specific theories that have been put forward. One such content-specific theory suggests that testing may shift attentional focus onto new material by reducing proactive interference of the old material (Dang, Yang, Che, et al., 2021; Dang, Yang, & Chen, 2021). After testing, the previously studied material becomes associated with both a study and test phase, whereas the new learning block would only be associated with a study phase. This results in a separate memory context being created for the previous material, reducing possible interference of the old material and shifting more attentional resources towards the new material (Karpicke et al., 2014; Yang et al., 2018). Critically, under this perspective, if the old and new learning material were in dissimilar domains, there should be no proactive interference during the new learning phase, so testing would have minimal to no impact. Therefore, this framework argues that similarity of content domain should moderate the strength of the testing effect, being stronger when the tested material and the to-be-learned material are in similar domains and weaker when they are in dissimilar domains.

A second group of content-specific theories suggests that testing leads to better encoding processes during subsequent learning (Chan et al., 2018). During a test, learners practice retrieving previously learned information, which increases the likelihood that those items are spontaneously retrieved or activated during the subsequent learning phase. This activation facilitation then increases the chance that new information is bound to old information in a more integrated mental model of the material, thus increasing learning of the new information (Jacoby et al., 2015). Based on this, two lessons in dissimilar domains should not be as easily integrated as two lessons in similar domains. Like the proactive interference explanation, if the forward testing effect is due to activation facilitation, content domain would moderate the strength of the forward testing effect and would be stronger for lessons in similar domains.

A third possibility, based on the Cognitive Theory of Multimedia Learning framework, suggests that testing may act as a signal indicating the more important types of information in a lesson (i.e., the signaling principle; Mayer, 2021, Van Gog, 2021), thereby facilitating efficient distribution of attentional resources to the types of key information during the new learning phase. In this case, the forward testing effect may be less dependent on content domain and more dependent on content structure, as the previous test signals when or how to-be–tested information is presented. Therefore, under this framework, content structure should moderate the strength of the testing effect, being stronger when the lessons have similar structures than when lessons have dissimilar structures.

Finally, a fourth possible explanation is that better encoding during the new learning phase is due to learners extracting an encoding strategy during the previous learning phase. If the encoding strategy was effective during the first learning block and reinforced by high performance on the test, learners could then transfer this strategy to the second block. However, this would only be effective if the learning strategy is useful in the new learning phase, which may depend on how the lesson is structured. This is similar to research on learning by analogy, in which finding an effective solution to a problem (in this case, an encoding strategy) transfers to a novel situation (the second learning block; Holyoak, 2012). For example, using an encoding strategy to learn English-Spanish word pairs may not be very useful during a calculus lesson, but could be useful in learning medical terminology. Similar to the CATML explanation, based on the encoding strategy theory, the forward testing effect should be stronger when the two lessons have similar content structures than dissimilar structures.

2 | CURRENT STUDY

Given the substantial number of theories regarding the forward testing effect, our goal here was to adjudicate between the two main classes of theories—content-general and content-specific theories—by varying the content domain and structure of the lessons. Specifically, our study was a 2 (Test vs. Study) × 2 (Similar vs. Different Domain) × 2 (Similar vs. Different Structure) fully mixed design. Although some previous work has suggested that the testing effect does occur if the tested and to-be-learned materials are from different content domains (e.g., star formation and lightning formation—Yue et al., 2015) and with different content structures (e.g., between declarative knowledge learning and category learning—Yang et al., 2019), no work to date has explicitly examined the magnitude of the forward testing effect as a direct function of the similarity of the content domain or lesson structure. The forward testing effect thus far has only been examined after relatively short time periods between the previously learned and to-be-learned lesson (e.g., 20 min) within the same day. However, it is unknown whether this effect is similar after a one-day delay or longer. As such, we implemented a one-day delay between the first and second lesson. If the forward testing effect occurs as a content-general process, the magnitude of the effect should be equivalent regardless of whether the domain/structure of the two lessons are similar or dissimilar. However, if the forward testing effect is content-specific, there should be an interaction between testing and content domain and/or structure similarity.

3 | METHOD

3.1 | Participants

Based on a previous meta-analyses that reported a large forward testing effect size for prose and video stimuli of $g = 0.77$ (Chan et al., 2018), a power analysis was conducted and revealed a necessary sample size of 11 per group, or 88 total for all conditions. Because the effects of content and structure similarity have not previously been estimated, the target sample size was doubled to be conservative. A total of 189 participants were recruited from the undergraduate Psychology participant pool at the University of Wisconsin—Madison. Participants who scored lower than chance on the multiple-choice questions (<2.5, $N = 10$) or scored higher than 2.5 standard deviations above the average total score (>13.65, $N = 3$) were excluded from analyses, leaving a total of 176 participants (46.9% male, 53.1% female, $M_{age} = 18.82$, SD = 1.14) in the final analyses. Participants were awarded extra credit for participating.

3.2 | Materials

3.2.1 | Video lessons

Four sets of two video lessons were used for each combination of the Domain and Structure conditions as the learning materials for this study1.[i] In the Same Domain conditions, both lessons were from the same academic domain (astronomy), while in the Different Domain conditions, the two lessons were from different academic domains (computer science, astronomy). These particular lesson topics were chosen as they were unlikely to have been studied in high school and would be challenging for undergraduates (i.e., that individuals would be unlikely to score well on a test on the topic in the absence of any instruction). In the Similar Structure conditions, both lessons contained declarative knowledge of a procedure (i.e., facts presented in a stepwise manner). Meanwhile, the first lessons used in the Different Structure conditions contained a series of discrete facts about a topic, while the second lesson presented declarative knowledge in a procedure. Critically, the second lesson in all four conditions were the same (a procedural astronomy lesson on the formation of the universe), which allowed for direct comparisons in performance.

1. Similar Domain, Similar Structure condition: the first lesson was a procedural astronomy lesson on the formation and death of white dwarfs.
2. Similar Domain, Different Structure condition: the first lesson was a factual astronomy lesson on the discovery and properties of dark matter.
3. Different Domain, Similar Structure condition: the first lesson was a procedural computer science lesson on training a neural network.
4. Different Domain, Different Structure condition: the first lesson was a factual computer science lesson on Alan Turing's life and contribution to computer science.

3.2.2 | Learning outcome measures

After the first lesson, those in the test conditions were asked to answer a cued recall prompt (e.g., "Please describe the discovery of dark matter and its properties in as much detail as possible.") and 10 multiple-choice questions. The free recall questions after the first test were scored according to a rubric and were out of either six points (Dark Matter, Neural Networks) or seven points (White Dwarves, Alan Turing). In the Study conditions, participants were asked to restudy the material by reading a summary of the content in the lesson. After the second lesson, all participants were then given the same cued recall prompt (i.e., "Please describe the steps of how the universe was formed in as much detail as possible."), which was scored according to a rubric and was out of a possible six points, and 10 multiple choice questions, each worth one point.

3.2.3 | Individual difference measures

A pre-lesson questionnaire collected demographic information and prior knowledge (e.g., participants were asked to list grades received in previous university courses in science, technology, engineering, and mathematics (STEM) classes, and specifically computer science/statistics and astronomy/physics). Additionally, the Motivated Strategies for Learning Questionnaire (MLSQ) was used to assess learners' domain-general strategies and effort (Duncan & McKeachie, 2005). Metacognitive strategies during learning were measured using the Awareness of Independent Learning Inventory (AILI; Meijer et al., 2013).

3.3 | Procedure

Figure 1 shows an overview of the study procedure. Participants signed up for a 2-day online study. They were first randomly assigned to view one of four lessons, which was either in the same domain

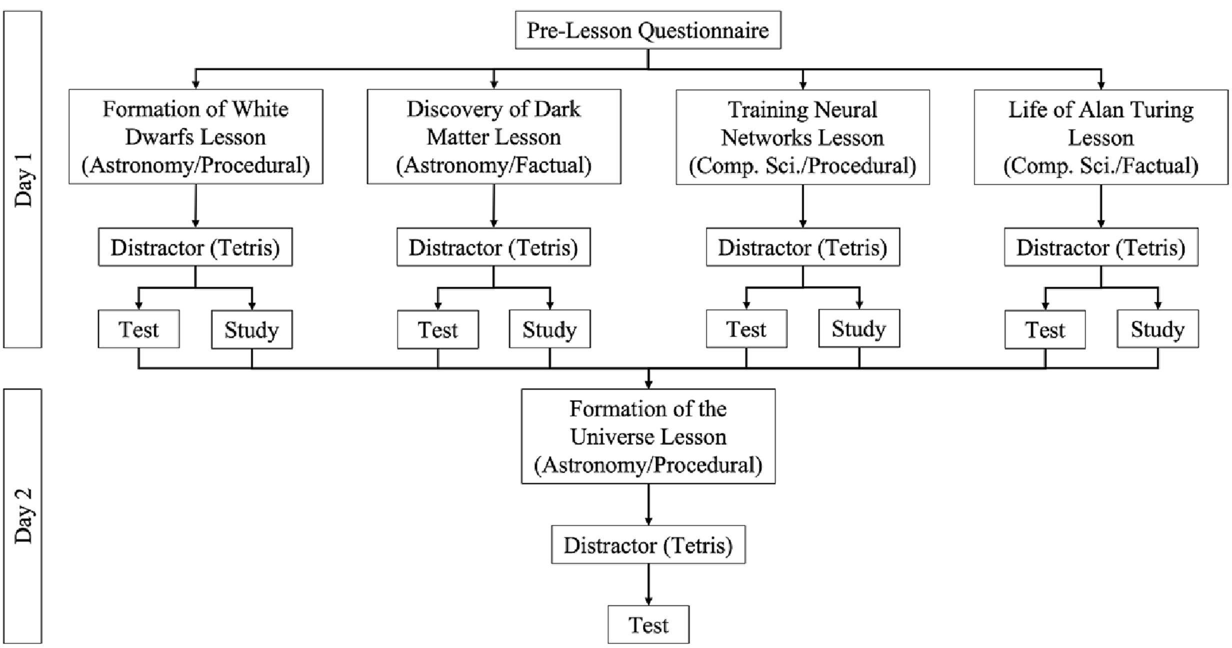

FIGURE 1 Flowchart of the study's procedure.

(astronomy) or different domain (computer science) as the subsequent lesson (astronomy) and was presented in either a similar structure (procedural) or dissimilar structure (factual) as the subsequent lesson (procedural). On the first day, participants completed the pre-lesson questionnaire, viewed their randomly assigned video lesson, and then completed a distractor task, which was a web-based version of Tetris, similar to the distractor task used in Yue et al. (2015). Then, those in the test conditions were asked to complete a quiz on the material they viewed, while those in the study condition were given a written summary of the material they viewed. On Day 2, all participants viewed the same video lesson (procedural astronomy lesson), completed a distractor task, and then were tested on the material from the second lesson.

This overall design allows us to examine main effects and interactions (e.g., a main effect of testing condition, in which individuals in test conditions outperform individuals in study conditions, would be consistent with an overall forward testing effect) across the eight conditions. However, we further a priori planned to test the extremes of the "same-different" dichotomy by comparing performance specifically in those participants whose lessons were the same in both domain and structure with those participants whose lessons were different in both domain and structure. Though theories underlying the forward testing effect do not make explicit predictions about main effects of domain or structure, content-specific theories would be supported if there was an interaction between testing and domain and/or structure, while content-general theories would be supported if there was significant evidence of the no interaction between testing and domain and/or structure. The procedures for this study were approved by the University of Wisconsin—Madison Institutional Review Board (IRB). Data for this study are available by emailing the corresponding author. This study was not preregistered.

4 | RESULTS

For those in the testing conditions, total test scores from the free recall and multiple-choice items on the day 1 test were calculated. A one-way ANOVA showed no differences in day 1 test scores across the testing conditions, $F(3, 96) = 1.93$, $p = .129$. Means and standard deviations for each condition are shown in Table 1.

Total test scores, including the free recall and multiple-choice scores after the second lesson, were calculated. A priori predictions of interactions between testing, domain, and structure were also tested. Because the content-general theories of the forward testing effect predict no interactions, Bayes' Factors (BF_{10}) were also included to examine evidence for the null effect. According to typical convention, BF scores between 1–3, 3–10, and 10–30 would indicate anecdotal, moderate, and strong evidence, respectively, in favor of the alternative hypothesis, while scores between 1/3–1, 1/10–1/3, and 1/30–1/10 would indicate anecdotal, moderate, and strong evidence in favor of the null hypothesis.

To test the three main effects, a 2 (Test vs. Study) × 2 (Similar vs. Different Domain) × 2 (Similar vs. Different Structure) ANCOVA was run on total test scores, with overall grade point average (GPA) and number of STEM classes used as covariates as these variables significantly correlated with test performance ($p = .037$ and $p = .048$, respectively). Consistent with previous results, a forward testing effect was found, $F(1,163) = 4.42$, $p = .037$, $n_p^2 = 0.03$, $BF_{10} = 1.31$, as shown in Figure 2 and Table 1. Those who were tested after the first lesson ($M = 6.85$, $SE = 0.30$) scored significantly higher than those who restudied the lesson ($M = 5.96$, $SE = 0.29$). Neither the main effect of domain nor of structure were significant ($F(1,163) = 0.16$, $p = .686$, $BF_{10} = 0.17$, and $F(1,163) = 0.01$, $p = .937$, $BF_{10} = 0.17$, respectively).

The interaction between testing and domain was not significant, $F(1,163) = 1.43$, $p = .233$, $BF_{10} = 0.220$. The interaction between testing and structure was also not significant, $F(1, 163) = 0.28$, $p = .868$, $BF_{10} = 0.206$. Finally, the three-way interaction between testing, domain, and structure was not significant $F(1,163) = .002$, $p = .962$, $BF_{10} = 0.035$. Taken together, the lack of interactions between testing and domain/structure and moderate evidence for the null models suggest minimal impact of content on the strength of the testing effect.

To further examine whether strength of the testing effect differed across lessons, a 2 (Test vs. Study) × 4 (Same domain/Same structure, Same domain/Different structure, Different domain/same structure, Different domain/different structure) ANCOVA was run on total test scores, with overall GPA and number of STEM classes used

TABLE 1 Means and standard deviations of test scores after the first and second lessons for each condition.

Condition			Day 1 test score		Day 2 test score	
Day 1 lesson	Domain	Structure	M	SD	M	SD
Test	Same	Same	7.13	3.05	7.10	3.11
Test	Same	Different	8.26	3.48	7.32	2.81
Test	Different	Same	6.30	2.47	6.68	2.71
Test	Different	Different	6.84	2.32	6.28	3.39
Study	Same	Same	–	–	5.66	2.16
Study	Same	Different	–	–	5.71	2.73
Study	Different	Same	–	–	6.14	2.39
Study	Different	Different	–	–	6.16	2.42

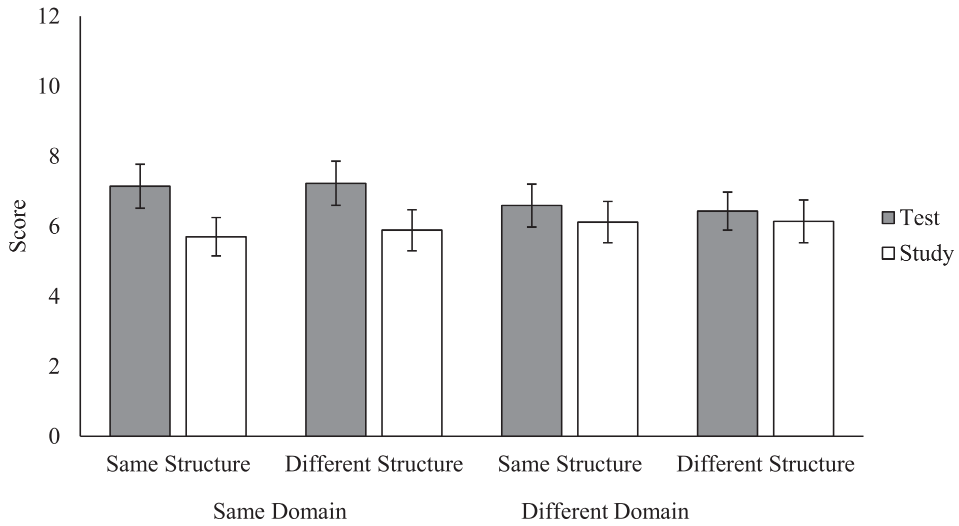

FIGURE 2 Total quiz scores by condition. The model includes GPA and the number of STEM classes taken as covariates. Error bars represent standard error.

as covariates. Here, an interaction would indicate that the strength of the testing effect differed between two or more of the lessons. However, the interaction was not significant, $F(1,163) = 0.493$, $p = .668$. A Bayes factor of 0.043 of this interaction also provides moderate evidence for the null model, suggesting that the forward testing effect is similar across the four types of lessons regardless of lesson domain and structure similarity.

Relationships between learning strategies, metacognitive strategies, and overall learning performance were also explored. Bivariate correlations were run between total quiz scores, five subscales of the MSLQ (self-efficacy, intrinsic value, cognitive strategy use, self-regulation, and test anxiety), and three subscales of the AILI (metacognitive knowledge, regulation, and response). Only one significant relationship emerged; total quiz score was significantly associated with metacognitive knowledge, $r(174) = 0.16$, $p = .039$. No other correlations were significant.

5 | DISCUSSION

The results from this study suggest that the forward testing effect occurs regardless of the similarity of the tested and to-be-learned material. The testing effects seems to be robust across lessons, and content domain similarity seems to have negligible effects on the forward testing effect. The results from this study show support for content-general theories over the content-specific theories. Furthermore, our work extends previous research by demonstrating the presence of a forward-testing effect after a one-day delay (i.e., rather than the very short delays utilized in much of the research to-date).

5.1 | Theoretical implications

Both the release-of-PI and activation facilitation theories predicted that the forward testing effect would be weaker when the lesson pairs differed in content domain, while the signaling and strategy change theories predicted it would be weaker when lesson pairs that differed in content structure. However, participants performed relatively similarly across all types of lessons. Thus, these theories cannot fully explain the forward testing effect. One caveat may be that there is weak support for the release-of-PI and activation facilitation theories as the difference between testing and study groups for lesson pairs in similar domains were numerically larger, though non-significant, than lesson pairs in dissimilar domains. Indeed, using other types of learning materials, particularly list learning, researchers have found support for these theories (Yang, Zhao, et al., 2021). Given the directional (though not statistical) effects seen here, future research may want to examine further probe these effects. Regardless, the effect of content domain does not seem to outweigh the effect of the forward testing effect overall.

Instead, the results from this study are more supportive of content-general theories of the forward testing effect. Testing may generally increase attention, metacognitive strategies, or motivation and effort during the subsequent lesson, thus increasing learning and learning outcomes. One parsimonious explanation may be that a general test expectancy effect increases motivation, which in turn also increases attention and metacognitive strategies. Previous work has indeed shown that expecting a later test improved subsequent learning outcomes (e.g., Agarwal & Roediger 3rd., 2011; Eitel & Kühl, 2015; Middlebrooks et al., 2017; Weinstein et al., 2014). Future work should test whether testing indeed increases test expectancy, attention, metacognition, or motivation, and whether changes in those mediate the forward testing effect.

Lastly, while the results here showed that there are significant relationships between general metacognitive traits and overall learning outcomes, it may also be important to examine whether and how testing changes these over time.

5.2 | Practical implications

In line with previous research, the forward testing effect was found across different types of learning material, which may have practical implications in the classroom (Yang et al., 2019; Yue et al., 2015). As testing seems to benefit future learning overall, one application may be to implement interim tests both within the same class and across classes in different academic domains. However, the generalizability of the forward testing effect in this study may be limited to other contextual factors associated with the online video lessons used in this study. For example, as all the lessons were produced by the same source, there may have been other contextual similarities that facilitated future learning that were not examined in this study, such as the pacing of information or relatability towards or trust of the instructor. Thus, it would be important for future research to examine this effect in real-world classrooms contexts.

One limitation of this study is its small sample size per group. Although the power analysis suggested that the testing effect should be detectable, it did not take into account the small effects of content similarity. Another limitation is that it used specific sets of lessons. It may be possible that these results do not generalize to other types of learning material in other domains or containing other structures. In addition, the learning material consisted of video lessons of academic-type material that required retention of declarative knowledge. Future work should determine whether the forward testing effect also occurs for other types of lessons or types of knowledge, such as category learning or motor learning.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available at https://doi.org/10.17605/OSF.IO/Y2VDN.

ORCID

Jocelyn Parong https://orcid.org/0000-0001-7076-2535

ENDNOTE

[i] All videos are publicly available on the CrashCourse YouTube channel (https://www.youtube.com/channel/UCX6b17PVsYBQ0ip5gyeme-Q).

REFERENCES

Agarwal, P. K., & Roediger, H. L., 3rd. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory*, *19*, 836–852.

Boustani, S., & Shanks, D. R. (2022). Heterogeneity and publication bias in research on test-potentiated new learning. *Collabra: Psychology*, *8*(1), 31996.

Bufe, J., & Aslan, A. (2018). Desirable difficulties in spatial learning: Testing enhances subsequent learning of spatial information. *Frontiers in Psychology*, *9*, 1701.

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*(11), 1111–1146.

Dang, X., Yang, C., Che, M., Chen, Y., & Yu, X. (2021). Developmental trajectory of the forward testing effect: The role of reset-of-encoding. *The European Journal of Developmental Psychology*, *19*, 869–884.

Dang, X., Yang, C., & Chen, Y. (2021). Age difference in the forward testing effect: The roles of strategy change and release from proactive interference. *Cognitive Development*, *59*(June), 101079.

Duncan, T. G., & McKeachie, W. J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist*, *40*(2), 117–128.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280.

Eitel, A., & Kühl, T. (2015). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, *11*, 107–121.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 234–259). Oxford University Press.

Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1282–1297.

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, *22*(3), 305–318.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 61, pp. 237–284). Elsevier Academic Press.

Lee, H. S., & Ahn, D. (2018). Testing prepares students to learn better: The forward effect of testing in category learning. *Journal of Educational Psychology*, *110*(2), 203–217.

Mayer, R. (2021). Cognitive theory of multimedia learning. In R. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (pp. 57–72). Cambridge University Press.

Meijer, J., Sleegers, P., Elshout-Mohr, M., van Daalen-Kapteijns, M., Meeus, W., & Tempelaar, D. (2013). The development of a questionnaire on metacognition for students in higher education. *Educational Research*, *55*(1), 31–52.

Middlebrooks, C. D., Murayama, K., & Castel, A. D. (2017). Test expectancy and memory for important information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 972–985.

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*, 286.

Pastötter, B., Engel, M., & Frings, C. (2018). The forward effect of testing: Behavioral evidence for the reset-of-encoding hypothesis using serial position analysis. *Frontiers in Psychology*, *9*, 1197.

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 287–297.

Risko, E. F., Buchanan, D., Medimorec, S., & Kingstone, A. (2013). Everyday attention: Mind wandering and computer use during lectures. *Computers & Education*, *68*, 275–283.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210.

Roets, A., Van Hiel, A., & Cornelis, I. (2006). The dimensional structure of the need for cognitive closure scale: Relationships with "seizing" and "freezing" processes. *Social Cognition*, *24*(1), 22–45.

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, *73*, 99–115.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*(16), 6313–6317.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1392–1399.

Tempel, T., & Frings, C. (2019). Testing enhances motor practice. *Memory & Cognition*, *47*(7), 1270–1283.

Van Gog, T. (2021). The signaling (or cueing) principle in multimedia learning. In R. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (pp. 221–230). Cambridge University Press.

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long- term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1039–1048.

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face–name learning. *Psychonomic Bulletin & Review*, *18*(3), 518–523.

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, *18*(6), 1140–1147.

Yang, C., Chew, S.-J., Sun, B., & Shanks, D. R. (2019). The forward effects of testing transfer to different domains of learning. *Journal of Educational Psychology*, *111*(5), 809–826.

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*, 399–435.

Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied*, *23*(3), 263–277.

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *NPJ Science of Learning*, *3*, 8.

Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 485–492.

Yang, C., Zhao, W., Luo, L., Sun, B., Potts, R., & Shanks, D. R. (2021). Testing potential mechanisms underlying test-potentiated new learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*, 1127–1143.

Yue, C. L., Soderstrom, N. C., & Bjork, E. L. (2015). Partial testing can potentiate learning of tested and untested material from multimedia lessons. *Journal of Educational Psychology*, *107*(4), 991–1005.

How to cite this article: Parong, J., & Green, C. S. (2023). The forward testing effect after a 1-day delay across dissimilar video lessons. *Applied Cognitive Psychology*, 1–8. https://doi.org/10.1002/acp.4101