Journal of Formative Design in Learning (2018) 2:129–143
https://doi.org/10.1007/s41686-018-0025-5

A Formative Design Examining the Effects of Elaboration and Question Strategy with Video Instruction in Medical Education

Julie Bridges[1] · Jill Stefaniak[2] · John Baaki[3]

Published online: 26 November 2018
© Association for Educational Communications & Technology 2018

Abstract
The purpose of this formative design study was to investigate the effects of instructor-created elaborations, learner-created elaborations, and adjunct questions on medical student learning outcomes in an asynchronous learning environment, using pre-recorded video over the course of a semester. The study also investigated the effects of instructor-created elaborations, learner-created elaborations, or adjunct questions on perceived cognitive load. The effect of learning strategy on quality of elaboration was also investigated. Results showed no significant difference in learning outcomes or cognitive load or quality of elaboration, but a post hoc analysis revealed a significant difference in intrinsic cognitive load for students who used generative strategies while having no gain in learning outcomes. Formative design data suggest that additional attention is needed on the learning environment when implementing generative learning strategies.

Keywords Generative learning · Medical education · Cognitive load · Formative design

Modes of delivery such as lecture capture or asynchronous video offer the convenience of virtual class attendance and the additional benefit of having class recordings available for review at any time. Design of recorded video instruction may not include consideration of instructional strategies to promote learning. A lack of consideration of instructional strategies and effective message design within the affordance of the delivery technology can interfere with learning outcomes (Anglin and Morrison 2000; Grabowski 2004). New technology tools for developing and delivering instruction arrive frequently and require instructor research and attention. Just as a patient will come to a physician asking about a new drug treatment they viewed in an advertisement, instructors frequently come to an instructional designer requesting help with the newest tool. Consideration of instructional strategies to use within any tool may help both instructors and instructional designers choose the best tools to suit the strategies appropriate for the learners.

Medical education is looking for ways to decrease lecture-based instruction and move toward independent and engaged learning. Traditionally, medical student education during the pre-clinical years involved lecture-based instruction of basic science facts, but now accrediting agencies are promoting independent learning, engaged learning, and situated learning (LCME 2016). Presentation of basic science facts may take place in a flipped environment. Flipped learning includes short visual presentations that are viewed in video format by individual learners before attending a learning event such as case-based learning or team-based learning (Tune et al. 2018; McLaughlin et al. 2014).

Inclusion of instructional strategies in these flipped videos may yield improved outcomes. Medical education is striving to embrace these new learning methods that support collaborative learning as a means to move away from traditional didactic lectures (Pluta et al. 2013). First and second year medical learners credit the learning environment and level of support from faculty as the main reason for burnout (Dyrbye et al. 2009). Perhaps a way to increase student satisfaction and decrease burnout is to change the learning environment and offer recorded video when appropriate. An increasing number of medical schools are utilizing recorded lectures to supplement learning of material in the first and second years of medical school (AAMC 2013).

✉ Jill Stefaniak
jill.stefaniak@uga.edu

[1] Eastern Virginia Medical School, PO Box 1980, Norfolk, VA 23501-1980, USA

[2] University of Georgia, 221 Rivers Crossing, 850 College Station Rd., Athens, GA 30605, USA

[3] Old Dominion University, 43rd and Hampton Boulevard #4125, Norfolk, VA 23529, USA

An increasing number of medical schools are utilizing recorded lectures to supplement learner learning of foundational knowledge in the first and second years of medical school (AAMC 2013). Flipped learning utilizes short videos, either gleaned from lecture videos or created anew, to prepare learners with foundational knowledge before participating in team-based learning (TBL) activities (Prober and Khan 2013). Generative learning strategy research has shown a positive effect on learning outcomes when compared to no strategy use (Anderson and Reder 1979; Divesta and Peverly 1984; Johnsey et al. 1992; Moreno and Mayer 2005). The inclusion of the generative learning strategy of elaborations in learning with video is hypothesized to improve learning over no elaborations and learning from video alone. In a study utilizing flipped learning in a pharmacology class, the first offering of videos did not include any generative learning and the instructors found the learners to be unprepared for class. In the second offering, the videos were be accompanied by questions (a form of generative learning) to ensure the learners are acquiring the necessary knowledge before coming to class and participating in collaborative learning activities (Mclaughlin et al. 2014). In a second year pathology course utilizing TBL, lower performing learners benefited from working with better prepared learners (Koles et al. 2005). The more prepared learners have a better experience with collaborative learning because they have not synthesized the to-be learned material. The LCME is encouraging medical schools to use collaborative learning experiences. The utilization of generative learning strategies with recorded video may be one way to prepare learners for collaborative learning experiences.

Purpose of the Study

The purpose of this formative design study was to investigate the effects of the generative strategies of instructor-created elaborations, learner-created elaborations, and adjunct questions on learning outcomes in an asynchronous learning environment, using pre-recorded video. The literature suggests that the use of generative strategies will positively affect learning outcomes but is unclear regarding the optimal use of generative strategies to support learning of foundational medical science principles at the application level. It was hypothesized that groups using the strategies of instructor-created elaborations, learner-created elaborations, or adjunct questions will perform higher than a video only/control group.

A Generative Approach to Instruction

When designing instruction, the first step in designing effective instruction is to focus on the instructional objective and clearly define the objective in terms of learner performance (Mayer 2010). After defining what the learner will do with the knowledge, a natural way to connect instruction to learning may be to have students do something with the newly acquired knowledge. Generative learning strategies work by creating relationships and meaning in restructuring information (Grabowski 2004; Jonassen 1988; Wittrock 1992). Including these generative learning strategies may help to accomplish the goal of having learners do something with the newly acquired knowledge from recorded video. Knowledge of how the technology supports an instructional strategy and the affordances it can give for learning may yield improved results (Koehler and Mishra 2014). Using technology for instruction that consumes both the development and delivery time of faculty and the misuse of mental resources of learners is both inefficient and unproductive, even if the desired learning outcomes are achieved (Molenda 2009). Careful consideration of strategies within the use of technology may improve learning outcomes.

Studies have shown an increase in learning outcomes when using generative learning strategies such as note-taking, highlighting concept mapping, and answering adjunct questions, but the results of other strategies such as elaboration are less clear (Grabowski 2004). Medical education utilizes elaborations in the form of short cases. Elaborations involve learners either creating examples themselves or studying provided examples. Even when the objectives are aligned and the integration of previous knowledge exists, no difference between types of elaboration may exist (Johnsey et al. 1992). Another way to increase engagement is to include questions in the instruction.

Adjunct questions are questions inserted into instruction to facilitate learning (Linder and Rickards 1985) and are considered a form of generative learning as long as the adjunct question requires the learner to integrate prior knowledge (Grabowski 2004). This study considered the strategies of elaborations (learner created and instructor created) and adjunct questions.

Adjunct Questions

Martyn (2009) directs designers to use questions as the first of seven principles of good practice to make recorded lectures interactive. Questions need to be crafted according to evidence-based principles (Andre 1979; Grabowski 2004; Rickards 1979). Research on adjunct questions has typically been conducted using text-based instructional materials either in print or on the computer (Rothkopf 1965; Rothkopf and Billington 1974; Rothkopf and Bisbicos 1967; Frase 1968; Mayer 1975; Rickards and Divesta 1974). Another study (Bow et al. 2013) showed in non-significant improvement in learning outcomes when students created flashcard questions for medical school.

The generative power of adjunct questions generally depends on the type, frequency, position, if they promote organizational activities, and if are written at an appropriate level of learning (Andre 1979; Grabowski 2004; Rickards 1979). Generally, application level questions inserted periodically into the material will yield the best results (Andre 1979; Rickards and Denner 1978; Rickards 1979). Application-level questions require learners to choose from a set of examples (Krathwohl 2002). Designers are also encouraged to use real-world examples (Martyn 2009). After the initial consideration to use adjunct questions (Martyn 2009), a designer may also consider using elaborations.

Elaborations

Studies where learners create examples themselves have shown improvement of retention of verbal information (Arkes and Freedman 1984; Divesta and Peverly 1984; Dooling and Christiansen 1977). Studies where learners use provided examples created by the subject matter expert instructors have shown improved learning for generated syntax examples but did not improve conceptual learning (Reder et al. 1986). Instructor-created examples have been shown to facilitate learning under specialized conditions such as high prior knowledge (Rothkopf and Billington 1974) or when the elaborations were more precise and clarified objectives (Stein and Bransford 1979). The efficacy of the elaborations is mainly dependent on alignment with objectives (Mayer 2010; Stein and Bransford 1979) and integration of previous knowledge (Grabowski 2004). Generative learning strategy research has shown a positive effect on learning outcomes when compared to no strategy use (Anderson and Reder 1979; Divesta and Peverly 1984; Johnsey et al. 1992; Mayer 1980; Moreno and Mayer 2005; Reder 1979; Stein and Bransford 1979).

Instructor-Created Elaborations Instructor-created elaborations are a type of example created by the instructor and/or subject matter expert. These created examples provide learners with the opportunity to study a completely correct worked out example of the topic at hand. Instructor-created elaborations do not always increase learning and in some cases have been found to harm retention of verbal information from text because the focus was not on the targeted learning objectives (Allwood et al. 1982; Reder and Anderson 1980). Instructor-created elaborations on computer skills improved learning for generated syntax examples but did not improve conceptual learning (Reder et al. 1986). Targeting the objectives to be learned, Stein and Bransford (1979) demonstrated that focusing the instructor-created elaborations on the objectives was helpful in the retention of verbal information. Instructor-created elaborations have been shown to facilitate learning under specialized conditions such as high prior knowledge (Rothkopf and Billington 1974) or when the elaborations were more precise and clarified objectives (Stein and Bransford 1979).

Learner-Created Elaborations Learner-created elaborations require learners to add their own understanding of a topic to create an example. Learner-created elaborations have shown to improve retention of verbal information (Arkes and Freedman 1984; DiVesta and Peverly 1984; Dooling and Christiansen 1977). Mayer (1980) investigated learner-created elaborations in problem solving specifically with computer programming skills. Mayer's (1980) research investigated two types of learner-created elaborations and found both to be effective in helping learners apply concepts and solve problems. One group compared new material to a model and the other group compared new pieces of information to each other. Both treatment groups utilized learner-created elaborations by including a form of self-explanation to scaffold the learner during elaboration creation. The results showed that learner-created elaborations were effective in low-ability and low prior knowledge subjects. Learner-created elaborations worked in this case because the learner engaged in two cognitive processes (Mayer 1980). First, the learner searched for prior knowledge and actively related their prior knowledge to the new information presented. The second cognitive process was the addition of self-explanation. The elaboration group outperformed the control group Mayer 1980. Self-explanation works by requiring the learner to answer "why" questions or explain the process they used to arrive at an answer (Chi et al. 1994). Spontaneous self-explanations do not always come easily (Renkl 1997) and eliciting self-explanations improves learning outcomes and learners with low prior knowledge received more benefit from elicitation (Renkl 1999). Prompting learners to self-explain each step of a probability problem improved performance (Renkl et al. 2002). The literature suggests that self-explanation works but is unclear as to what form of elaboration works best. Using elaborations may yield improved learning over no elaborations but the literature is still unclear as to what type of elaboration is ideal. Johnsey et al. (1992) found no difference between learner-created elaborations and instructor-created elaborations. Self-explanations may be successfully used with worked examples (Sweller 2010) given the learners have sufficient cognitive resources to self-explain. The conditions governing cognitive resources involved in elaborations and self-explanation might be found in cognitive load theory (Sweller et al. 2011). If the cognitive load does not exceed the capacity of the learner's resources, then the learner will have the ability to create examples and self-explain during the process.

Self-Explanation Effect

Self-explanation can improve learning by encouraging learners to generate explanations (Bielaczyc et al. 1995; McNamara 2004; Renkl 1997). According to Clark et al. (2006, p. 190), self-explanation is a mental dialog that the learner has when studying worked examples that helps the learner develop schema. More knowledgeable learners who have the resources to self-explain will benefit from the process. Novice learners may not have the resources available for the organizing and linking principle to function (Sweller et al. 2011). Self-explaining enhanced knowledge acquisition for eighth grade learners studying the circulatory system (Chi et al. 1994). Spontaneous self-explanations do not always come easily (Renkl 1997) and eliciting self-explanations improves learning outcomes and learners with low prior knowledge received more benefit from elicitation (Renkl et al. 1998). Prompting learners in computer environments to self-explain each step of a probability problem (Atkinson et al. 2003) improved performance. Self-explanations can be successfully used with worked examples (Sweller et al. 2011) given the learners have sufficient resources to self-explain. Eliciting self-explanations helps learners with low prior knowledge. These learners are in jeopardy of not having sufficient resources for germane processing (Sweller et al. 2011) and training with elaborations improves learning outcomes (Johnsey et al. 1992). Studying or creating examples may require more cognitive resources from the learner than answering adjunct questions. If the process of elaboration creation imposes a greater expenditure of cognitive resources on the learner, the increase in essential processing may exceed available cognitive resources. In order to design elaborations with self-explanation, an understanding of cognitive load theory is required.

Generative Strategies and Cognitive Load

Using the principles of cognitive load theory, instructional designers can manage the limitations of working memory and maximize the extensive capabilities of long-term memory. Strategies that encourage strategic usage of germane processing and reduction of extraneous cognitive load were of particular interest in this study because the content to be learned contains complex interrelated concepts and appears to have high intrinsic cognitive load. Strategies that encouraged germane processing include worked examples (instructor-created elaborations). The worked example effect discourages the means-end style of problem solving, reduces cognitive load, and facilitates schema construction. An instructor-created elaboration in anatomy is a case that contains all the steps involved in solving a problem regarding an area of the body. Worked examples focus the learner on the steps of the problem solution. Identifying important features of the worked examples for the learner to attend will improve the learner's experience with the worked example (Anderson et al. 1990). The important features in a worked example or instructor-created elaboration can be attended to through the use of self-explanation. In the domains of algebra, statistics, geometry, and programming, using worked examples is beneficial to learning outcomes (Cooper and Sweller 1987; Sweller and Cooper 1985; Paas 1992; Paas and van Merriënboer 1994; Trafton and Reiser 1993). Worked examples are effective because learners view worked examples as a primary source of learning material (Lieberman 1986; Pirolli 1991; Segal and Ahmad 1993). Some disadvantages to using worked examples are a lack of training with learner problem solving tasks, and stereotyping of solution patterns may be reasons to include other forms of instructional procedures such as completion problems (Sweller et al. 1998). The worked example effect directly applies to the instructor-created elaborations that are a completed example from the primary source of the instructor. Learner-created elaborations may require learners to utilize a greater amount of cognitive resources and may exceed available resources for germane processing. Training to use strategies may be necessary in order for learners to successfully complete the strategies during instruction, and to ensure extraneous cognitive load is not imposed while learning to use a strategy (Fiorella and Mayer 2015).

Training to Use Strategies

Selecting appropriate strategies such as learner-created examples or instructor-created examples depends on prior knowledge of the learner and the nature of the material to be learned (Wittrock 1990). Fiorella and Mayer (2015) summarized empirical studies on the strategies using summarizing, concept mapping, drawing, imaging, self-testing, self-explaining, and enacting while reporting on the varying types of training and varying lengths of time used for training. According to Wittrock's (1990) generative learning theory, the selection of strategy should be based on prior knowledge and nature of material, the type of material, and the time for training. Fiorella and Mayer (2015) recommended verbal strategies such as summarizing, self-testing, self-explaining, and teaching for non-spatial material and strategies such as mapping, drawing, imagining, and enacting for spatial material. The training time spent for these strategies varied based on type of material and on learner prior knowledge. Training learners to use generative strategies is needed for all strategies and should be based on type of material to be learned and learner prior level of knowledge (Fiorella and Mayer 2015). Previous empirical studies for undergraduate level and above learners ranged from 19 to 30 min (Azevedo and Cromley 2004; Catrambone and Yuasa 2006; Stark et al. 2002).

A formative evaluation of four medical student study groups was conducted as the first of a series of interventions to examine the effects of elaboration and question strategy with video instruction and the implications on medical students' cognitive load.

Formative Research Questions

The following research questions guided this study:

1. What is the effect of generative learning strategy (i.e., instructor-created elaborations, learner-created elaborations, adjunct questions, and video-only) on learning outcomes?
2. What is the effect of generative learning strategy (i.e., instructor-created elaborations, learner-created elaborations, adjunct questions, video-only) on cognitive load?
3. What effect does the generative learning strategy have on elaboration quality?

Method

Research Design

This study used a quasi-experimental design approach to examine the research questions. Participants were randomly assigned to one of the four generative learning strategy groups (learner-created elaborations, instructor-created elaborations, adjunct questions, and video-only/control group). The interventions for the four strategy groups were designed as an initial micro-cycle for a multi-phased design project (McKenney and Reeves 2012). This initial micro-cycle placed emphasis on carrying out the three main processes of exploring solutions, mapping solutions, and constructing prototypes during the design and construction phases of this formative design project. The independent variable used in the study was generative learning strategy operationalized as instructor-created elaborations (ICE), learner-created elaborations (LCE), adjunct questions treatments (AQ), and a video-only/control group (VO).

The learner-created elaboration group created elaborations (or examples) of simplified medical cases after watching a video in anatomy. The instructor-created groups studied elaborations created by the subject matter expert after watching a video in anatomy (see Appendix for examples). The adjunct question group answered application-level questions after watching a video in anatomy. The video-only/control group watched a video in anatomy.

Participants and Setting

Participants for this study included first year undergraduate medical students enrolled in their first semester anatomy course at an urban medical school in the Mid-Atlantic region. The course was delivered via a traditional lecture/laboratory format. The lectures were recorded in a previous year and were utilized for asynchronous recorded lectures instead of face-to-face lectures. The course also provided supplemental recorded lectures available to all learners. The number of medical students enrolled in a class is 150. The course was given in one section with one main course director and multiple instructors. Exclusion criteria for this study consisted of learners who previously attended a medical school, or repeated the class.

Measures

Learning Outcomes Instructor-created multiple-choice test questions measured learning outcomes in a post-test after each module. The learning outcomes were measured by the post-test score ranging from 0 to 100. Each of the module post-tests included 15 knowledge-level questions along with 5 application-level questions (Krathwohl 2002). Questions with an item difficulty between .60 and 1 p value were accepted and utilized. The post-test measured the effects of generative strategies on learning outcomes after each of the three lessons. Reliability of the post-test instrument was assessed using a KR-20 reliability coefficient.

Cognitive Load The modified NASA-TLX survey with four dimensions of effort, mental demand, performance, and frustration level, measured on a scale of 1–100, examined the effects on cognitive load for each of the experimental conditions after each module. The NASA-TLX survey measured workload while studying elaborations provided by the instructor, creating elaborations, or answering application-level adjunct questions. The survey is a modified version of the NASA-TLX assessment tool (Hart and Staveland 1988). The NASA-TLX is a subjective tool that allows users to rate their perceived level of mental demand on a continuous 100-point scale. The entire tool measures mental demand, physical demand, temporal demand, overall performance, frustration level, and effort. The section measuring mental demand was utilized for this study. The NASA-TLX instrument has been used in over 500 studies and was developed by the Human Performance Group at NASA. All learners answered questions regarding their cognitive load after each lesson. The results of the cognitive load questions were analyzed to determine if a relationship exists between perceived workload and post-test scores. The questions used for percieved workload were formatted as a continuous scale rating from 1 to 100. The selection of the NASA-TLX instrument was based on the desire to measure the three dimensions of cognitive load.

Elaboration Quality Quality of elaborations is a factor that only applied to the leaner-created elaboration group and the instructor-created group because the other two groups did not complete self-reflections. Participants in these groups (instructor-created and learner-created) had to complete a written response to the self-reflection questions for each part of the examples they either created or studied. Each question response was judged by the same rubric for both the learner-created group and the instructor-created group of elaborations. The instructor-created and learner-created elaboration groups generated responses to elaboration and self-reflection prompts in their treatment groups after each module. The quality of responses was measured with a four-dimension, three-level rubric that was used to give feedback to the instructor-created and learner-created groups. The dimensions of case presentation, origin, insertion and action, imaging or tests, and differentials were measured at three levels. The rubric was used to give feedback to the learner on the quality of their responses in the learner-created and the instructor-created groups for each module.

Procedures

Learners were randomly assigned to one of the four generative learning strategy groups labeled instructor-created elaborations, learner-created elaborations, adjunct questions, and video-only.

Instructor-Created Elaborations Learners attended regular classes and watched a supplemental instructional video in anatomy. To train learners to do self-reflections, instructions were provided for the learners to review before studying an instructor-created elaboration. Learners then studied instructor-created elaborations. These expert elaborations were created by the instructor to help learners get more information about a certain topic such as how origin, insertion, and action of a muscle could be affected by a condition. Learners were required to self-explain, in writing, why each part of the instructor examples was appropriate and submit their answers in a course management system (CMS). An example of excellent and poor answers to self-reflection questions was provided to learners in the beginning of each topic for training in how to study elaborations.

Learner-Created Elaborations Learners attended regular classes and watched a supplemental instructional video in anatomy. To train learners to create elaborations and self-reflections, instructions were provided for the learners to study before creating their elaborations. Learners then created their own examples or elaborations about a certain topic such as how origin, insertion, and action of a muscle could be affected by a condition using an outline to prompt for the required parts of the elaboration. Learners reflected on the parts of their examples as they created them to explain why they consider their answers to be a good example. The learner-created elaborations were submitted in the CMS. Feedback was given to the learners after each submission in the form of a rubric. The rubric was completed by a trained teaching assistant who had mastered the content in anatomy and returned to the learner before the next topic is presented.

Adjunct Questions Learners attended regular classes and watched a supplemental instructional video in anatomy. Questions were answered after every 10–15 min of video in a natural break on the topic. Questions were presented to learners in the CMS, and learners were encouraged to go back and review the video to write down the answers to the questions. Answers were presented to and collected from the learner in test management software.

Video Only Learners attend regular classes and watched a supplemental instructional video in anatomy. Learners were also prompted to watch a recording of the regular lecture given in class to control for time on task.

Results

Analysis of Learning Outcomes

Learning outcomes were analyzed by four groups to evaluate the differences between adjunct questions, learner-created examples, instructor-created examples, and video-only treatment groups (Table 1). Table 2 represents the means and standard deviations of learning outcomes for the four groups.

The results of the analysis revealed no significant difference in test performance. An analyses of variance (ANOVA) was performed for quiz 1. No significant differences were between groups on learning outcomes, $F(3,29) = .295$, $p = .829$, $\eta^2 = .03$. An ANOVA was performed for quiz 2. No significant differences were between groups on learning outcomes, $F(3,29) = .852$, $p = .477$, $\eta^2 = .08$. An ANOVA was performed for quiz 3. No significant differences were found between the four groups on learning outcomes, $F(3,29) = .224$, $p = .879$, $\eta^2 = .02$.

Analysis of Cognitive Load

For all trials, participants in the create examples group had the highest mean of all groups while the participants in the video-only group had the lowest mean for all trials. Table 3

Table 1 Research design

Groups	Measures	
Adjunct questions	Learning outcomes	Cognitive load
Learner-created examples	Learning outcomes	Cognitive load
Instructor-created examples	Learning outcomes	Cognitive load
Video only	Learning outcomes	Cognitive load

Table 2 Means and standard deviations of learning outcomes by treatment group

		Quiz (trial)					
		1		2		3	
Group	*n*	*M*	SD	*M*	SD	*M*	SD
Adjunct questions	13	60.39	15.06	61.54	17.96	67.69	21.18
Learner-created examples	5	63.00	16.05	56.00	17.10	66.00	13.42
Instructor-created examples	8	63.75	14.58	68.75	8.35	59.38	16.35
Video only	7	59.29	7.87	59.27	10.58	62.86	22.33

Scores range from 0 to 100 for all items

represents the means and standard deviations of total cognitive load for the four groups by trial.

Participants in the learner-created examples group reported the highest levels of overall mental effort in trial 1 ($M = 58.33$, $SD = 11.16$), followed by adjunct question ($M = 56.66$, $SD = 12.31$), and followed by instructor-created examples group ($M = 53.40$, $SD = 10.47$). Participants in the video-only group reported the lowest overall mental effort ($M = 45.11$, $SD = 10.91$).

Participants in the learner-created examples group reported the highest levels of overall mental effort in trial 2 ($M = 65.50$, $SD = 16.58$), followed by adjunct question ($M = 53.56$, $SD = 14.45$), and followed by instructor-created examples group ($M = 51.31$, $SD = 10.86$). Participants in the video-only group reported the lowest overall mental effort ($M = 43.10$, $SD = 13.34$). Participants in the learner-created examples group reported the highest levels of overall mental effort in trial 3 ($M = 64.33$, $SD = 15.39$), followed by instructor-created examples group ($M = 55.31$, $SD = 9.28$), followed by adjunct question ($M = 55.12$ $SD = 10.67$), and participants in the video-only group reported the lowest overall mental effort ($M = 43.69$, $SD = 15.67$).

Total cognitive load measures were analyzed by four groups to evaluate the differences between adjunct questions, learner-created example, instructor-created example, and video-only treatment groups. The results of the analysis revealed no significant difference in cognitive load. A repeated ANOVA was performed. No significant differences were found between the four groups on total cognitive load, $F(3,29) = 2.10$, $p = .122$, $\eta^2 = .22$.

Table 4 shows the means and standard deviations by dimension of demand in cognitive load by trial. A repeated measures ANOVA was used to measure the differences in total demand by groups. The results show no significant differences in the dimensions of demand by the four groups of adjunct questions, learner examples, instructor examples, and video-only $F(3,29) = 2.79$, $p = .058$, $\eta^2 = .22$.

Table 5 shows the means and standard deviations by dimension of effort in cognitive load by trial. A repeated measures ANOVA was used to measure the differences in total effort by groups. The results show no significant differences in the dimensions of effort by the four groups of adjunct questions, learner-created examples, instructor-created examples, and video-only $F(3,29) = .889$, $p = .458$, $\eta^2 = .084$.

Table 5 shows the means and standard deviations by dimension of frustration in cognitive load by trial. A repeated measures ANOVA was used to measure the differences in total frustration by groups. The results show no significant differences in the dimensions of frustration by the four groups of adjunct questions, learner-created examples, instructor-created examples, and video only $F(3,29) = .475$, $p = .702$, $\eta^2 = .047$ (Table 6).

Analysis of Elaboration Quality

The quality of elaboration was measured with a rubric. The means for the two groups, study instructor-created and learner-created groups, are presented in Table 7.

Table 3 Means and standard deviations of total cognitive load by four treatment groups

		Total load measurement (trial)					
		1		2		3	
Group	*n*	*M*	SD	*M*	SD	*M*	SD
Adjunct questions	13	56.66	12.31	53.56	14.45	55.12	10.67
Learner-created examples	5	58.33	11.16	65.50	16.58	64.33	15.39
Instructor-created examples	8	53.40	10.47	51.35	10.86	55.31	9.28
Video only	7	45.11	10.91	43.10	13.34	43.69	15.67

Scores range from 0 to 100 for all items

Table 4 Means and standard deviations of demand dimension of cognitive load by treatment group

		Trial					
		1		2		3	
Group	*n*	*M*	SD	*M*	SD	*M*	SD
Adjunct questions	13	60.19	18.10	58.85	18.14	55.77	18.04
Learner-created examples	5	65.50	15.45	70.00	20.92	68.00	17.17
Instructor-created examples	8	54.36	23.05	49.06	12.95	55.31	16.50
Video only	7	48.93	22.82	40.00	18.92	35.00	20.32

Scores range from 0 to 100 for all items

Means for the study examples were consistently higher than the learner-created examples group for case grades for each trial. A repeated ANOVA was performed to examine the differences between the two groups of study instructor-created examples and learner-created examples group on the case grades for each group. No significant difference was found between groups on case grades, $F(2,10) = 1.268$, $p = .323$, $\eta^2 = .06$.

Post Hoc Analysis

The results of this study did not support most of the previous literature conclusions in and may be in part because of the nature of the medical student population. Medical students are highly motivated and have high prior science knowledge. In order to further understand these results, a post hoc analysis was performed to understand the results related to cognitive load measures. A question considered in the analysis was whether cognitive load differed significantly for the groups utilizing the strategies from those that did not.

To further investigate cognitive load measures and the effect generative strategies have on cognitive load, a post hoc analysis was performed. Data were analyzed by learners using generative strategies vs. non-generative strategies by combining all generative strategy treatment groups against the video-only (control) group. An ANOVA was performed. A significant difference was found between groups on total cognitive load, $F(1,31) = .609$, $p = .019$, $\eta^2 = .16$. The results of the analysis revealed a significant difference in total cognitive load. An ANOVA was performed to analyze differences in the three dimensions of cognitive load as measured by the NASA-TLX. The three dimensions are demand, effort, and frustration. Cognitive load measures were significantly different between two groups on the dimension of demand, $F(1,31) = 5.79$, $p = .022$, $\eta^2 = .157$. No significant difference was found on the dimensions of effort or frustration.

Discussion

Participants who utilized generative learning strategies had no difference in learning outcomes as compared to participants who did not utilize generative learning strategies. These results do not support Wittrock's (1990, 1991, 1992) theory of generative processing regarding the engagement of generative strategies to prompt learners in making new connections between prior and new knowledge. The results do not support the research results of Stein and Bransford (1979), DiVesta and Peverly (1984), and Johnsey et al. (1992) that utilizing elaborations would positively affect learning outcomes. One potential reason may be the type of learners who participated in these studies. The three previously mentioned studies involved undergraduate learners or adult learners utilizing elaborations as a generative learning strategy and this population may be very different from medical student learners.

Table 5 Means and standard deviations of effort of cognitive load by treatment group

		Trial					
		1		2		3	
Group	*n*	*M*	SD	*M*	SD	*M*	SD
Adjunct questions	13	54.23	26.83	46.92	23.76	51.54	22.86
Learner-created examples	5	52.50	23.63	62.50	28.72	50.00	23.09
Instructor-created examples	8	46.44	18.80	49.44	29.00	58.89	22.61
Video only	7	38.57	21.16	35.71	19.88	37.14	30.94

Scores range from 0 to 100 for all items

Table 6 Means and standard deviations of dimension of frustration of cognitive load by treatment group

		Trial					
		1		2		3	
Group	*n*	*M*	SD	*M*	SD	*M*	SD
Adjunct questions	13	41.92	30.86	44.62	30.71	36.15	30.14
Learner-created examples	5	36.00	24.03	40.00	33.91	44.00	37.81
Instructor-created examples	8	53.75	35.43	43.13	26.58	56.25	32.92
Video only	7	38.57	28.97	30.00	21.60	34.29	28.20

Scores range from 0 to 100 for all items

Population

Medical students may be very different in terms of knowledge and motivation than undergraduate learners who participated in the previously cited studies. Post-baccalaureate students like medical students are high-achieving students that may have superior skills in choosing learning strategies. Medical students may have practiced many strategies in learning science material as evidenced by their good grade point averages and above average Medical College Admission Test (MCAT) scores.

Adjunct Questions

The results do not support the hypothesis that using application-level adjunct questions during an online video presentation would increase learning outcomes (Andre 1979; Rickards and Denner 1978; Rickards 1979). Other studies using a medical school population found no differences in learning outcomes. The results did support the study on flashcard use in medical education that showed in non-significant improvement in learning outcomes when students created flashcard questions (Bow et al. 2013). In the aforementioned study, an increase in overall exam scores was observed but because demographic data was not collected, a measurement of differences by groups was not possible. Students who were skilled in question creation may have been the ones to self-select to participate in the group to use the strategy. The researchers did not measure cognitive load to see whether self-selection of strategy use impacted cognitive load. Because of the pairing of students to answer questions in the Mclaughlin et al. (2014) study, the results of this study cannot be directly compared. The study found a significant difference in learning outcomes using generative strategies; however, students had the choice to work in pairs to answer questions in a flipped learning environment. The synergy of pairing in answering questions could have made the difference in increasing learning outcomes.

Self-Explanations

Additionally, results of this study do not support Chi et al.'s (1994) argument that self-reflection is superior to only interacting with instructional materials, specifically, self-reflection during the use of elaborations. One possible reason for this may lie in the knowledge and motivation of learners participating in this study. The higher knowledge and higher motivation of the learners participating may occlude any benefits of self-reflection because the learners are already engaged and motivated to learn the material. Multiple studies (Bielaczyc et al. 1995; McNamara 2004; Renkl 1997) found benefits of self-explanation in many contexts but all learners had low prior knowledge. Chi et al. (1994) found that eighth grade learners studying the circulatory system benefitted from self-explanation, but the population in this study was post-baccalaureate medical students with extensive prior knowledge of science.

One explanation from the literature is that low prior knowledge learners have fewer prior knowledge connections to make with new knowledge and the self-explanation prompts

Table 7 Means and standard deviations of rubric grades by examples treatment group

		Quiz (trial)					
		1		2		3	
Group	*n*	*M*	SD	*M*	SD	*M*	SD
Learner-created examples	4.47	5	73.00	10.37	79.00	6.52	77.00
Instructor-created examples	8	82.50	16.69	88.13	13.08	89.37	14.00

Scores range from 0 to 100 for all items

provide opportunity for learners to make new connections (Renkle 1997; Renkl et al. 1998). Additionally, Renkl et al. (1998) found that learners with low prior knowledge received more benefit from elicitation. The learners in this study were all medical students with a high prior knowledge in biology, chemistry, physics, and organic chemistry.

The characteristics of a typical medical student may yield some additional explanation about the learning outcomes and cognitive load outcomes in this study. Medical students are generally highly motivated, have high prior knowledge in science, are time-impoverished, and may have developed their own preferences for use of generative strategies (Nair et al. 2013; Kusurkar et al. 2013). The average number of hours medical students spend in class or studying is 10 h per day (AAMC 2016), and they have a high level of stress as measured by the Perceived Stress Scale (AAMC 2016).

Cognitive Load

The addition of generative strategies did not affect learning outcomes in any generative group and an investigation of cognitive load may yield an explanation for these findings. The original research question investigated the difference in cognitive load for four groups of instructor-created examples, learner-created examples, adjunct questions, and video only.

Measures of Cognitive Load Cognitive load measures were affected by the use of generative strategies in this study. The measurement of overall cognitive load from the NASA-TLX was affected by the use of generative learning strategies. The specific measurement of demand was also affected by generative strategies at a significant level but no other dimension of cognitive load was significantly affected. The overall cognitive load measure is parsed out with six questions to measure the areas of demand, effort, and frustration. Specifically, the measurement of demand was significantly affected by the use of generative learning strategies. The measure of demand may be a measurement of intrinsic cognitive load as proposed by Gerjets et al. (2006). The demand measure increased without providing the benefit of increased learning outcomes. One reason for this could be that the layering of any prescribed strategy on top of the learning material may be unnecessarily increasing the demand, or intrinsic cognitive load. These results point to an increase in intrinsic cognitive load from the use of generative learning strategies. Leppink et al. (2013) reported that the measures of intrinsic load and germane load may not be linear. If a learning task is easy, the explanation and instructions for the task may not contribute to learning. Leppink et al. (2013) argue that if the learning experience was too complex for a learner, germane load capacity may be limited. More research may be needed to clarify the measures of intrinsic load and germane load.

Generative Learning Increased Cognitive Load In this study, cognitive load was reported in the moderate range from 45 to 58 on a 100-point scale so the task may not have been difficult enough even though it was significantly different for the learners using generative learning strategies. Specifically, an increase in cognitive load was measured in the dimension of demand that significantly increased for the group of learners using generative strategies. While learning outcomes were not negatively affected, the outcomes were also not positively affected by the use of generative strategies and may be because the subject of anatomy is both verbal and spatial in nature. In this study, the addition of generative learning strategies did not benefit the learner. The addition of generative learning strategies increased intrinsic load for students and consumed resources that would otherwise be available for germane processing. One explanation for this finding may lie in Fiorella and Mayer's (2015) recommendation to use verbal strategies (summarizing, self, and explanation) for non-complex or non-spatial material and spatial generative strategies (mapping, drawing) for teaching complex spatial concepts. The discipline of anatomy has both verbal and spatial components and may benefit from using spatial generative strategies such as creating flashcards.

Mental Effort Measurement in Self-Explanation The task of self-explanation and generation of examples was thought to decrease in difficulty with practice. The overall mental effort was not significantly different in trials 1 and 2 but significantly differed in trial 3. The mean cognitive load increased in the generative group for trial 3, thus not supporting the idea that elaboration over repeated measures decreases in difficulty and may decrease cognitive load. A repeated measure using the elaboration (examples) may not be necessary for this population, especially creating elaborations in the form of cases because the majority of instruction in medical school is in the form of cases. The dimension of demand was significant in trial 3 for learners using generative strategies. The level of overall mental effort significantly increased in trial 3 and did not decrease as expected. An explanation for overall cognitive load not decreasing as the trials advanced, for the study examples group and create examples group, may be found in the finding that the generative strategies only increased the demand dimension of cognitive load and perhaps did not make more germane resources available for processing the learning material.

Mental Effort in Strategy Use The dimension of effort as a subscale of cognitive load measure was found to relate to germane processing by Gerjets et al. (2006). Higher means of effort were reported by learners in the generative learning group compared to the video-only group but no significant differences were found in any trial. Learners were attending to instructional material due to the higher reported mean, and germane processing was occurring, but not at a significantly different level in any trial. The increase of germane processing is desirable as long as resources are available for processing the learning material. In this study, the generative strategies were engaging enough to show an increase in the mean of effort in all generative categories over the video-only group, but not enough to reach statistical significance. The learning outcome means were higher in the generative groups, but not significantly higher. The learning outcome means were overall lower than expected for all groups. The learning material may not have been engaging enough to bring overall cognitive load to an optimal level. The relationship between intrinsic load and germane load may not be linear (Leppink 2013) because one learner may have low prior knowledge and interpret the task as demand, while another student may have high prior knowledge and interpret the task as effort. The relationship between prior knowledge and cognitive load may need some further research.

Possible Treatment Effects Typically, similar examinations in anatomy courses have a higher mean around 80%. One explanation for the higher mean with the regular assessments in anatomy could be the extensive laboratory experience that accompanies lecture. In this study, the instruction was all delivered with video and no opportunity to complete an actual laboratory. Although the video was a virtual dissection and review of the region, and the generative treatments were meant to increase attention directed to germane resources and processing, the treatments seemed only to increase intrinsic cognitive load and divert resources from potential germane processing. An examination of the measure of frustration on the NASA-TLX may yield some additional insight in terms of learning new concepts.

Extraneous Cognitive Load The NASA-TLX measure of the dimension of frustration was found to relate to extraneous cognitive load (Gerjets 2006); however, it was not found to be statistically different for any group. The means for frustration in the generative group was consistently lower in all three trials, but not at a level of significance. Adding generative strategies did not increase frustration for learners in the generative strategy group, but not significantly and not enough to impact learning outcomes. An explanation for this finding may be that adding generative strategies for medical students will slightly increase frustration and possible extraneous cognitive load by requiring a strategy in addition to one the student typically prefers to use.

Quality of Elaborations Quality of elaboration was examined to determine if the higher quality answers were related to those with higher measures of learning outcomes. There was no relationship between the two groups of study instructor-created examples (SE) and student-created examples (CE) group on the case grades for each group. One possible reason for this may lie in the type of learner that is highly motivated, higher previous science knowledge, and the nature of anatomy as both verbal and spatial in nature. The strategy of examples may not be the best strategy to use for learning anatomy and therefore may not provide any benefit to learning outcomes.

The quality of case grades did not improve over the course of three trials. The literature shows varying results regarding the utility of creating examples vs. studying examples. Instructor-created examples worked to improve improved learning for some cases including generated syntax examples but did not improve conceptual learning (Reder et al. 1986). Stein and Bransford (1979) showed that focusing the instructor-created elaborations on the objectives were helpful in the retention of verbal information. Instructor-created elaborations have been shown to facilitate learning under specialized conditions such as high prior knowledge (Rothkopf and Billington 1974). In this case, the learners did not demonstrate higher learning outcomes, but did report a higher intrinsic cognitive load from the treatment. This may be due to the learner interpreting the task as demand. The same type of varying results occurs in the literature regarding learner-created examples. Johnsey et al. (1992) found no difference between learner-created elaborations and instructor-created elaborations while Mayer's (1980) research found both to be effective in helping learners apply concepts and solve problems.

Questions to Consider

In creating instructional strategy experiences going forward, faculty could consider giving students choices for strategies to use in learning from recorded video rather than prescribing a single strategy to use for learning. A knowledge of learner preferences for tools may be obtained by surveying learners regarding the tools and strategies that previous cohorts of students used to learn. Utilizing the list of tools and strategies former students used to learn the material and including those tools and strategies as suggestions for student success may be way to obtain the positive results achieved in previous studies (Bow et al. 2013; Mclaughlin et al. 2014). For example, pairing students for questioning strategy use or self-selecting the tools to use individually may yield similar results.

Limitations to Study

A limitation of this formative design study was a relatively small number of participants ($n = 33$). Participant recruitment was hindered by the time required to participate in a repeated measures study requiring 2–3 h of time outside of the regular curriculum time. The required curriculum time is an average of 24 h per week, with the additional study hours on top of the required hours, not leaving much extra time to participate in this study. Incoming students have a varied background in anatomy depending on their undergraduate program. The anatomy guy supplemental videos are a source for learners to fill in knowledge gaps including practice test questions. One unanswered question that emerged from the study is the effect that the level of learner has on the measure of intrinsic cognitive load. An analysis of prior experience in anatomy from student transcripts in addition to overall science grade point average and MCAT admissions scores may reveal a relationship between prior knowledge and perception of cognitive load. The quality of cases did not improve over the course of three trials and one reason for this could be the familiarity that medical students already have with case creation (examples), during the course of their first semester, multiple courses, including a clinical course on doctoring, used cases in a standard format for instruction.

Considerations for Future Design Iterations

Future research for the area of strategy use based on learner pre-requisite knowledge holds great promise for medical education. The attributes for learners in medical education provide a rich environment for studying strategy use for a population of learners who are highly motivated, have high abilities, and have high previous knowledge.

The formative data gathered during this initial study will be utilized to further explore the relationship between elaboration strategies and cognitive load among medical students. The results of this initial design study identified challenges unique to medical education that warrant additional exploration. Medical students have been conditioned to learn in an environment under a schedule that violates cognitive load processes on a regular basis. As this has become commonplace, due to the nature of medical training in the USA, many students in this study responded negatively when attempts were made to alleviate cognitive load.

More learner analysis is needed in future research to analyze the qualities of medical school learners that make this population unique. Performance results from previous science undergraduate courses, specifically performance in a directly related subject such as anatomy, would have been useful for this study. The learner analysis may give insight for strategy use as pre-requisite knowledge may relate to how learners are applying knowledge and using strategies.

For years, designers were told that manipulation of intrinsic load was only possible by sequencing material and not possibly by adding strategies to the learning experience. In this study, adding strategies appeared to increase intrinsic cognitive load and future research is needed to investigate why this might happen for the population of leaners in medical education.

Using a new measurement tool may give insight into the increase in intrinsic load for high-knowledge populations. Additionally, future research may provide a more in-depth examination into the process of studying examples or creating examples that could reveal more information about the relationship between the effects of generative strategies and cognitive load using a mixed method think-aloud protocol. Additionally, a survey about preferred generative strategy use may yield insight into the issue of prescribing a strategy contrary to a preferred strategy that may increase intrinsic cognitive load. The think-aloud protocol may also yield insight into the issue of cognitive load while studying or creating examples. The learners with high prior knowledge may be able to articulate the perception of working against their preferred generative learning strategy. It would also be beneficial to have the learner complete the assigned instructional activities and articulate their understanding of the content through the use of a think aloud protocol to better understand the effects imposed on their cognitive load. Digging into the perception for details of why learners perceive material to be more demanding and may reveal valuable information.

Additional information about the actual relationship between the categories of the NASA-TLX and the category of intrinsic cognitive load may be needed in future iterations of this design study. There may be a difference in perception of intrinsic load by medical students that varies from the perception of intrinsic load for undergraduate learners. One way to measure this may be to have both undergraduate anatomy students at another university participate in the same treatment and compare their work aloud perceptions to the work aloud perceptions of admitted medical students.

Additionally, there may be differences in learning outcomes longitudinally over time. A possible future area of research would be to measure the retention of material after the basic science courses have concluded to see if strategy use affected long-term retention of material.

The high prior knowledge in science areas, learner-type admitted to medical school may enter the program with successful strategies already in place such as utilizing flash cards, concept maps, mnemonics, or drawings to learn anatomy material. An extension of this study could be to explore the relationship between choice of strategies and learning outcomes. For example, analyzing learning outcomes and cognitive load for learners who choose to use pre-made flash cards vs. learners who choose to create their own flash cards rather than randomly assigning learners to strategies. An analysis of strategy choice with cognitive load perception may reveal a higher level of metacognitive awareness within the population of medical students.

Conclusion

Medical education is looking for ways to decrease lecture-based instruction and move toward independent learning which is the self-assessment of learning needs; independent identification, analysis, and synthesis of relevant information; and appraisal of the credibility of information sources (LCME 2016). Traditionally, medical student education, during the pre-clinical years, involved rote memorization of basic science facts but now accrediting bodies are promoting independent learning, engaged learning, and situated learning. Methods such as case-based learning and the more structured team-based learning (TBL) encourage learners to take more responsibility for their learning and provide a more integrated approach to basic science learning rather than rote memorization of facts. Presentation of concepts may take place in a flipped environment. Flipped learning includes short visual (not text based) presentations that are viewed in video format by individual learners before attending a learning event such as case based or TBL. Medical education is struggling to embrace these new learning methods that support collaborative learning as a means to move away from traditional didactic lectures (Pluta et al. 2013). Part of the struggle stems from a lack of common definitions for types of learning and from how to appropriately engage medical students in the new types of learning (Pluta et al. 2013).

More research is necessary to parse out the effect that prescribing a generative strategy to a higher knowledge population such as medical students would have on learning outcomes. Using a think-aloud protocol may also reveal a difference of perception of demand among different populations of learners. Due to the subjectivity of cognitive load measures, there may be differences in perception based on the type of learner. This research has attempted to provide insight into the use of generative strategies for use with recorded presentations in an authentic medical anatomy classroom environment. Given the results of no significant impact to learning outcomes, and the significant increase in the demand measure of intrinsic cognitive load, these results may give the foundation to future research.

Compliance with Ethical Standards

Conflict of Interest The authors declare that there is no conflict of interest.

Appendix

Instructor-Created Elaboration Examples

Studying examples is a skill that requires some practice. In this study, you will study examples, created by your instructors, three different times. Print these directions out so you have them in front of you every time you study the instructor examples because you will answer some self-reflection questions and submit your answers in Blackboard. Self-reflection questions help you to think about *why* the instructor example is a good one. Each of the three times you turn in answers, you will receive feedback to help you improve your performance for the next time.

After watching the supplemental video, you will be given an example to study. To help you do the best job possible, here is a good example:

Step 1: Your instructor provided this text in bold font as an example to help you learn the anatomy of a region:

A 22 year-old female soccer player presents to the emergency department with a swollen right ankle after twisting it at a soccer game. The player reports falling on the ankle and hearing an audible "pop" after going up for a "header" with several other players. The ankle is swollen, bruised and tender to palpation. The player reported playing on the ankle for a few more minutes before leaving the game. The athletic trainer sent her to the emergency department after the game.

You will have to answer this question: Why is this a good case to help you learn the region? How will it help you to better identify the action, origin and insertion of muscles and/or the innervation and or the blood supply in that region?

Here is a good example of a Self-Reflection Answer: *This is a good case to learn the anatomy of the ankle because it is a common injury and could involve many structures that can be included in the differential diagnosis. Because this case could involve multiple muscles, identifying the origin, insertion, and actions of these muscles will help determine which one is injured. Muscles involved would depend on the mechanism of injury. If the patient inverted her ankle, then the muscles on the lateral aspect of her leg would be affected: peroneus longus and brevis, and peroneus tertius. Ligaments could also be involved (anterior talofibular ligament).*

OK, now you have seen a great example, let us look at a response that is not good. Notice the lack of detail in the answer. Essentially the information is correct, but there is not enough detail to show your complete understanding.

Here is a poor example of a Self-Reflection Answer: *This is a good case to learn the anatomy of the ankle because it is a common injury.*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Allwood, C. M., Wikstrom, T., & Reder, L. M. (1982). The effects of text structure on free recall: more support for summaries. *Poetics, 11*(2), 145–153.

Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 385–403). Hillsdale: Lawrence Erlbaum Associates.

Anderson, J. R., Boyle, C. F., Corbell, A., & Lewis, M. (1990). Cognitive 34ulticen and intelligent tutoring. *Artificial Intelligence, 42*(1), 7–49.

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research, 49*(2), 280–318.

Anglin, G. J., & Morrison, G. R. (2000). An analysis of distance education research: implications for the instructional technologist. *The Quarterly Review of Distance Education, 1*(3), 189–194.

Arkes, H. R., & Freedman, M. R. (1984). A demonstration of the costs and benefits of expertise in recognition memory. *Memory and Cognition, 12*(1), 84–89.

Association of American Medical Colleges, (2013). Organization of learner representatives. Lecture video recording and attendance. Retrieved from https://www.aamc.org/download/268838/data/lecture_video_recording_and_attendance.pdf. Accessed 12 Sept 2016.

Association of American Medical Colleges, (2016). About the MCAT exam. Retrieved from https://www.students-residents.aamc.org/applying-medical-school/taking-mcat-exam/about-mcat-exam/. Accessed 12 Sept 2016.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523–535.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221–252.

Bow, H. C., Dattilo, J. R., Jonas, A. M., & Lehmann, C. U. (2013). A crowdsourcing model for creating preclinical medical education study tools. *Academic Medicine, 88*(6), 766–770.

Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: the effects of examples and elaborations and active learning exercises. *Learning and Instruction, 60*(2), 139–153.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & La Vancher, C. (1994). Eliciting self-explanations improves understandings. *Cognitive Science, 18*(3), 439–477.

Clark, R. C., Nguyen, F., Sweller, J., & Baddeley, M. (2006). Efficiency in learning: evidence-based guidelines to manage cognitive load. *Performance Improvement, 45*(9), 46–47.

Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347–362.

DiVesta, F. J., & Peverly, S. T. (1984). The effects of encoding variability, processing activity, and rule-examples sequence on the transfer of conceptual rules. *Journal of Educaitonal Psychology, 76*(1), 108–119.

Dooling, D. J., & Christiansen, R. E. (1977). Episodic and semantic aspects of memory for prose. *Journal of Experimental Psychology: Human Learning and Memory, 3*(4), 428–436.

Dyrbye, L. N., Thomas, M. R., Harper, W., Massie, F., Power, D. V., Eacker, A., et al. (2009). The learning environment and medical student burnout: a multicenter study. *Medical Education, 43*(3), 274–282.

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: eight learning strategies that promote understanding*. New York: Cambridge University Press.

Frase, L. T. (1968). Some data concerning the mathemagenic hypothesis. *American Educational Research Journal, 5*, 181–189.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction, 16*(2), 104–121.

Grabowski, B. L. (2004). Generative learning contributions to the design of instruction and learning. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 719–743). New York: Lawrence Earlbaum.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North Holland.

Johnsey, A., Morrison, G. R., & Ross, S. M. (1992). Using elaboration strategies training in computer-based instruction to promote generative learning. *Contemporary Educational Psychology, 17*(2), 125–135.

Jonassen, D. H. (1988). *Instructional designs for microcomputer courseware*. Hillsdale: L. Erlbaum Associates.

Koehler, M. J., & Mishra, P. (2014). Introducing TPCK. In M. C. Herring, M. J. Koehler, & P. Mishra (Eds.), *Handbook of technological pedagogical content knowledge (TPCK) for educators* (pp. 13–40). New York: Routledge.

Koles, P., Nelson, S., Stolfi, A., Parmelee, D., & DeStephen, D. (2005). Active learning in a year 2 pathology curriculum. *Medical Education, 39*(10), 1045–1055.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory Into Practice, 41*(4), 212–218.

Kusurkar, R. A., Croiset, G., Galindo-Garré, F., & Ten Cate, O. (2013). Motivational profiles of medical students: association with study effort, academic performance and exhaustion. *BMC Medical Education, 13*, 87. https://doi.org/10.1186/1472-6920-13-87.

LCME, (2016). Liaison Committee on Medical Education. Retrieved from http://lcme.org/wp-content/uploads/filebase/standards/2017-18_Functions-and-Structure_2016-03-24.docx. Accessed 15 May 2017

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072.

Lieberman, H. (1986). An example-based environment for beginning programmers. *Instructional Science, 14*(3–4), 277–292.

Linder, R. W., & Rickards, J. P. (1985). Questions inserted into text: issues and implications. In D. Jonassen (Ed.), *Technology of text* (pp. 131–157). Englewood Cliffs: Educational Technology Publications.

Martyn, M. A. (2009). Engaging lecture capture: lights, camera… interaction. *Educause Quarterly Magazine, 32*(4). Retrieved from https://www.educause.edu/ero/article/engaging-lecture-capture-lightscamerainteraction. Accessed 2 May 2018.

Mayer, R. E. (1975). Forward transfer of different reading strategies evoked by test-like events in mathematics text. *Journal of Educational Psychology, 67*(2), 165–169.

Mayer, R. E. (1980). Elaboration techniques that increase the meaningfulness of technical text: an experimental test of the learning strategy hypothesis. *Journal of Educational Psychology, 72*(6), 770–784.

Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical Education, 44*(6), 543–549. https://doi.org/10.1111/j.1365-2923.2010.03624.x.

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research*. New York: Routledge.

McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., et al. (2014). The flipped classroom: a course redesign to foster learning and engagement in a health professions school. *Academic Medicine, 89*(2), 236–243.

McNamara, D. S. (2004). SERT: self-explanation reading training. *Discourse Processes, 38*(1), 1–30.

Molenda, M. (2009). Instructional technology must contribute to productivity. *Journal of Computing in Higher Education, 21*(1), 80–94.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection and interactivity in an agent-based multimedia game. *Journal of Educational Psychology, 97*(1), 117–128.

Nair, S. P., Shah, T., Seth, S., Pandit, N., & Shah, G. V. (2013). Case based learning: a method for better understanding of biochemistry in medical students. *Journal of Clinical and Diagnostic Research: JCDR, 7*(8), 1576.

Paas, F. (1992). Training strategies for attaining transfer of problem solving skill in statistics: a cognitive load approach. *Journal of Educational Psychology, 84*(4), 122–133.

Paas, F., & van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 51–71.

Pirolli, P. L. (1991). Effects of examples and their explanations in a lesson on recursion: a production system analysis. *Cognition and Instruction, 8*(3), 207–259.

Pluta, W. J., Richards, B. F., & Mutnick, A. (2013). PBL and beyond: trends in collaborative learning. *Teaching and Learning in Medicine, 25*(sup1), S9–S16.

Prober, C. G., & Khan, S. (2013). Medical education reimagined: a call to action. *Academic Medicine, 88*(10), 1407–1410.

Reder, L. M. (1979). The role of elaborations in memory for prose. *Cognitive Psychology, 11*(2), 221–234.

Reder, L. M., & Anderson, J. R. (1980). A partial resolution of the paradox of interference: the role of integrating knowledge. *Cognitive Psychology, 12*(4), 447–472.

Reder, L. M., Charney, D. H., & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. *Memory and Cognition, 14*(1), 64–78.

Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science, 21*(1), 1–29.

Renkl, A. (1999). Learning mathematics from worked-out examples: analyzing and fostering self-explanations. *European Journal of Psychology of Education, 14*(4), 477–488.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: the effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*(1), 90–108.

Renkl, A., Atkinson, R. K., Maier, U. J., & Staley, R. (2002). From example study to problem solving: smooth transitions help learning. *The Journal of Experimental Education, 70*(4), 293–315.

Rickards, J. P. (1979). Adjunct postquestions in text: a critical review of methods and processes. *Review of Educational Research, 49*(2), 181–196.

Rickards, J. P., & Denner, P. R. (1978). Inserted questions as aids to reading text. *Instructional Science, 7*(3), 313–346.

Rickards, J. P., & Divesta, F. J. (1974). Type and frequency of questions in processing textual material. *Journal of Educational Psychology, 66*(3), 354–362.

Rothkopf, E. Z. (1965). Some theoretical and experimental approaches to problems in written instruction. In J. D. Drumboltz (Ed.), *Learning and the educational process*. Chicago: Rand McNally.

Rothkopf, E. Z., & Billington, M. J. (1974). Indirect review and priming through questions. *Journal of Educational Psychology, 66*(5), 669–679.

Rothkopf, E. Z., & Bisbicos, E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology, 58*(1), 56–61.

Segal, J., & Ahmad, K. (1993). The role of examples in the teaching of programming languages. *Journal of Educational Computing Research, 9*(1), 115–129.

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction, 12*(1), 39–60. https://doi.org/10.1016/S0959-4752(01)00015-9.

Stein, B. S., & Bransford, J. D. (1979). Constraints on effective elaboration: effects of precision and subject generation. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 769–777.

Sweller, J. (2010). Cognitive load theory: recent theoretical advances. In Plass, Moreno, & Brunken (Eds.), *Cognitive load* (pp. 29–47). Cambridge: Cambridge University Press.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

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

Trafton, J. G., & Reiser, B. J. (1993). *The contribution of studying examples and solving problems to skill acquisition. Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 1017–1022). Hillsdale: Erlbaum.

Tune, J. D., Sturek, M., & Basile, D. P. (2018). Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology, 316–320. https://doi.org/10.1152/advan.00091.2013. Accessed 20 July 2018.

Wittrock, M. (1990). Generative processes of comprehension. *Educational Psychologist, 24*(4), 345–376.

Wittrock, M. (1991). Generative teaching of comprehension. *Elementary School Journal, 92*, 167–182.

Wittrock, M. (1992). Generative learning process of the brain. *Educational Psychologist, 27*(4), 531–541.