Received: 8 July 2019 | Revised: 30 March 2020 | Accepted: 1 April 2020
DOI: 10.1002/acp.3671

SPECIAL ISSUE ARTICLE

WILEY

When seeing what's wrong makes you right: The effect of erroneous examples on 3D diagram learning

Allison J. Jaeger[1] | Joanna A. Marzano[2] | Thomas F. Shipley[2]

[1]Department of Psychology, St. John's University, Jamaica, New York

[2]Department of Psychology, Temple University, Philadelphia, Pennsylvania

Correspondence
Allison J. Jaeger, Department of Psychology, St. John's University, 8000 Utopia Parkway, Marillac Hall, Jamaica, NY 11439.
Email: jaegerba@stjohns.edu

Funding information
Directorate for Social, Behavioral and Economic Sciences, Grant/Award Numbers: 10-41707, 1640800, 1839705

Summary

Comprehending 3D diagrams is critical for success in scientific practice and research demonstrates that understanding of 3D geology diagrams can be improved by making predictive sketches. In mathematics, explaining erroneous examples can support learning. This study combined these approaches to better understand how to effectively support 3D geologic diagram understanding. Participants generated sketches, explained erroneous example sketches, or copied and explained correct sketches. It was hypothesized that generating sketches or explaining erroneous cases would improve understanding, but an open question was whether these conditions would differ from each other. Explaining erroneous examples and sketching improved understanding whereas explaining correct sketches did not. Further, explaining erroneous examples was a more efficient strategy than sketching. These results indicate that erroneous examples can be effective for supporting 3D diagram comprehension and may be a practical substitute for some traditional sketching activities in the context of real classrooms where class time is limited.

KEYWORDS

diagrams, erroneous examples, science learning, sketching, spatial thinking

1 | INTRODUCTION

Learning in science, technology, engineering, and mathematics (STEM) is challenging because it requires understanding complex and abstract processes. Visual representations are often used to communicate and support understanding of many STEM concepts because the processes can be too small, large, fast, or slow to be directly perceived. Diagrams can be especially important tools for illustrating three-dimensional (3D) structures or relationships. For example, in chemistry, dash-wedge diagrams illustrate the spatial configurations of molecules at the atomic scale, and in geoscience, geologic block diagrams illustrate geologic structures that can span scales from centimeters to tens of kilometers. While diagrams are important tools in STEM and offer great potential as meaning-making resources, in practice they can be very challenging for students to understand (Roth, Bowen, & McGinn, 1999; Wu, Krajcik, & Soloway, 2001) especially when they convey critical 3D spatial relations (Cooper, Stieff, & DeSutter, 2017; Gagnier, Atit, Ormand, & Shipley, 2017; Kali & Orion, 1996). Failing to understand 3D diagrams could be a barrier to STEM success because understanding and generating these representations is key in scientific practice (Ainsworth, Prain, & Tytler, 2011).

The goal of the present study was to examine the effectiveness of different approaches for supporting 3D block diagram understanding in the domain of geology. In particular, the goal was to explore if explaining erroneous example sketches would improve geologic block diagram understanding and whether this improvement would be better than the improvements seen when copying and explaining correct sketches. An additional question was whether explaining erroneous examples would be as effective as generating sketches, a strategy often used to support 3D diagram understanding in geoscience.

2 | SKETCHING TO SUPPORT STEM LEARNING

External representations (such as diagrams, graphs, maps, and animations) are a central component of STEM education and are frequently encountered in textbooks and online sources (Cromley et al., 2013; Slough, McTigue, Kim, & Jennings, 2010). Instead of relying solely on words for communication and problem-solving, scientists often rely on a host of visualizations such as sketches, diagrams, photographs, and graphs to represent complex spatial concepts in their disciplines (Ainsworth et al., 2011; Newcombe & Stieff, 2012). Further, skilled scientists often self-generate visualizations to facilitate discovery and innovation (Nersessian, 1992). Being able to make sense of and create spatial representations is an essential component of spatial thinking, and visual literacy more generally (McTigue & Croix, 2010; National Research Council, 2012). Unfortunately, when visual literacy activities are implemented in the classroom, which is rare, they tend to focus students on interpreting preexisting representations that correctly depict spatial information rather than asking students to generate their own visual representations through sketching, or evaluate and interpret partially generated or partially incorrect representations (Ainsworth et al., 2011; Coleman, McTigue, & Smolkin, 2011). Thus, students may fail to develop spatial thinking skills, and in some cases disregard visual representations all together (Winn, 1994).

Because diagram comprehension and generation play a central role in becoming a successful scientist, there has been a surge of interest in investigating learner-generated diagramming for supporting STEM learning (e.g., Gagnier et al., 2017; Jaeger, Velazquez, Dawdanow, & Shipley, 2018; Leutner & Schmeck, 2014; Scheiter, Schleinschok, & Ainsworth, 2017; Schmeck, Mayer, Opfermann, Pfeiffer, & Leutner, 2014). Research has demonstrated that sketching can facilitate inferential reasoning and mental model development (Alesandrini, 1981; Gobert, 2005) and that sketches can be used for assessing knowledge (Johnson & Reynolds, 2005; Schwamborn, Mayer, Thillmann, Leopold, & Leutner, 2010). Results from these studies have reported medium to large effect sizes with Cohen's d values ranging from 0.50 to 0.90 (e.g., Leopold & Leutner, 2012; Leutner & Schmeck, 2014; Schmeck et al., 2014; Schwamborn et al., 2010). Sketching is especially helpful for learning in STEM because it more readily captures critical spatial information and aligns with the visual–spatial demands of STEM learning (Suwa & Tversky, 1997; Vosniadou & Brewer, 1992).

Sketching to support reasoning has a long history in STEM, especially in geology (Johnson & Reynolds, 2005; Turner & Libarkin, 2012). Geology deals with the Earth's physical structure and the processes that act on it, most of which are inaccessible to humans and take place beneath the Earth's surface. Thus, geologists often reason about unobservable 3D phenomena, and commonly use sketching to make predictions, record observations, and evaluate hypotheses. Sketching is also a pedagogical tool for reasoning about unobservable 3D structures. In the field, students sketch to reason about and *predict* the connections between outcrops (visible exposures of rock on the Earth's surface; Shipley, Tikoff, Ormand, & Manduca, 2013).

In geoscience, block diagrams illustrate how rock layers extend into the earth by depicting two sides and the top of a block that represents a 3D geologic structure (see Figure 1) and are foundational for teaching about the Earth's interior structure and deformation. Although common in introductory geology classes, students often err when making spatial inferences from block diagrams (Alles & Riggs, 2011; Atit, Gagnier, & Shipley, 2015; Kali & Orion, 1996; Ormand et al., 2014). Common errors include assuming the interior of the block is identical to the exterior face, failing to integrate information from multiple sides of the block, and failing to perceive any 3D relations at all. One of the most researched methods for teaching students to interpret and understand geologic block diagrams is to have them generate cross-sections of the blocks and then sketch the interior structure that is revealed by the cross-section. Problem-solving via sketching is commonly used in geoscience because it is an active spatial tool that can support learning and mental model updating via feedback and error identification (Alles & Riggs, 2011; Gagnier et al., 2017; Kali & Orion, 1996).

Gagnier et al. (2017) examined the efficacy of generating predictive sketches for improving reasoning about 3D geologic block diagrams. They suggested generating predictive sketches, sketches that require students to predict what the inside of a block diagram would look like if it were sliced into, could address students' spatial reasoning errors because it requires visualizing within- and between-object spatial relationships, fosters alignment, and allows one to see the spatial differences between one's own sketch and a correct sketch. In this study, students generated predictive sketches of block diagram cross-sections, then compared their sketches to correct sketches. Sketching led to improved 3D diagram understanding compared to only visualizing the cross-sections without sketching or copying correct sketches without making spatial predictions (medium effects were reported; d = 0.51–0.58). Gagnier et al. argued predictive sketching supported 3D diagram understanding because it required making inferences about spatial relationships that cannot be directly observed, and mapping those 3D inferences into a diagrammatic space. They also argued that sketching allows students to generate a representation they can compare to the correct answer, and then use this feedback to update their mental model.

3 | COGNITIVE RESOURCES NEEDED TO SKETCH

While sketching can support STEM learning (Cromley, Du, 2020), it may not help all students. Generating a sketch relies heavily on cognitive capacities like working memory and spatial reasoning and thus, may not be an effective learning strategy unless an individual possesses a certain threshold of such capacities. For example, visuospatial working memory has been shown to be critical for developing spatial mental models of route descriptions (Brunyé & Taylor, 2008a), especially when generating such representations demands a high degree of complex (i.e., 3D) mental imagery (Brunyé & Taylor, 2008b). Research has also shown that sketch quality correlates with spatial thinking

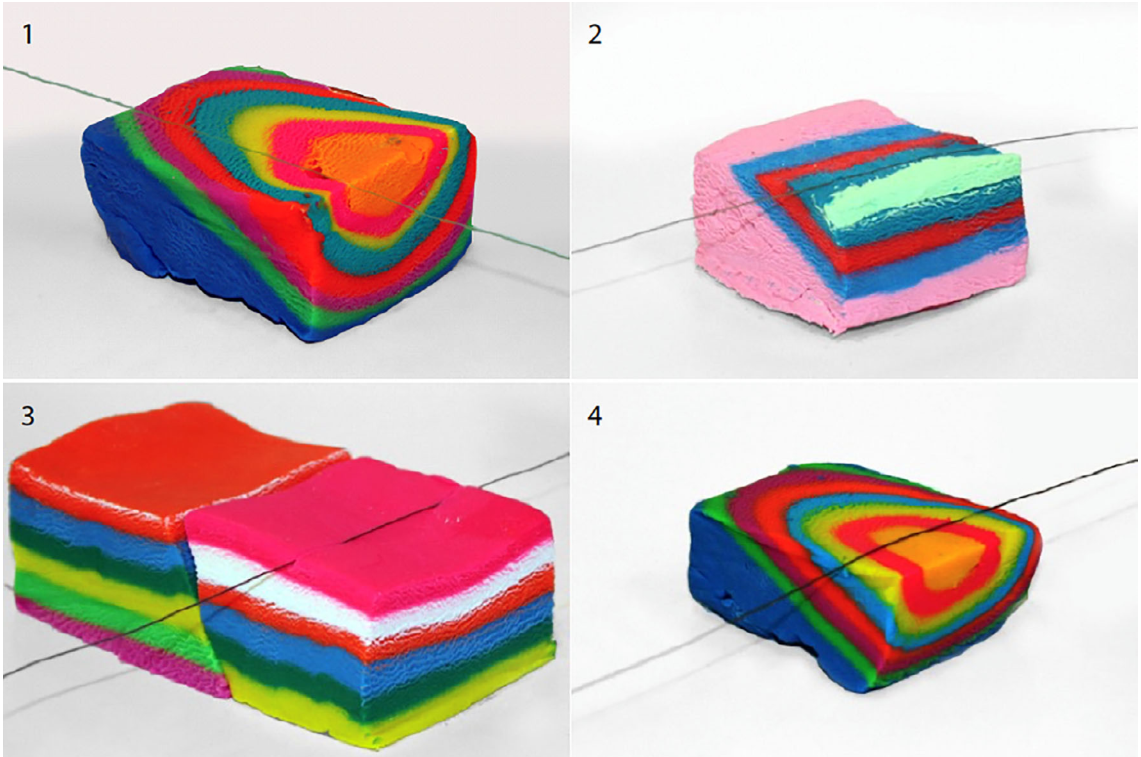


FIGURE 1 The four structures used as stimuli in order of presentation during the intervention: (1) deeply plunging fold; (2) dipping layers; (3) faulted horizontal layers; (4) gently plunging fold. The wire indicates where a cut will be made [Colour figure can be viewed at wileyonlinelibrary.com]

suggesting low-spatial individuals who struggle to develop spatial mental models may benefit less from sketching activities (Jaeger et al., 2018).

Furthermore, much of the research showing positive effects for sketching has done so in conditions where additional learner support is provided. For example, benefits have been found when students complete learning-strategy training (Leopold & Leutner, 2012), when they receive guidance about what elements to include in sketches (Alesandrini, 1981; Schmeck et al., 2014; Schwamborn et al., 2010), or when given partially complete sketches to fill in (Britton & Wandersee, 1997). These instructional supports help reduce the extraneous cognitive load generated during sketching and without support, sketching may consume a large amount of cognitive resources, leaving insufficient resources for integration and mental model building (Leutner & Schmeck, 2014). Providing corrective feedback or giving students the opportunity to compare their sketches to correct illustrations in order to correct spatial errors and update mental models has also improved learning (Fan, 2015; Van Meter & Garner, 2005).

Not only are sketching activities cognitively demanding, they are also time-consuming, both in terms of time required to complete sketching activities and time for providing constructive feedback (Garnier et al., 2017). Further, providing feedback to students regarding their sketches is challenging because they can be difficult to score reliably (e.g., White & Gunstone, 1992). Together, these practical burdens on instructors can result in delayed and less useful feedback for students (Dihoff, Brosvic, Epstein, & Cook, 2004; Kulik & Kulik, 1988), which is a critical component of effective sketching activities.

4 | WORKED EXAMPLES AND ERRONEOUS EXAMPLES

Worked examples may be an effective alternative learning strategy to sketching because they impose fewer cognitive and practical constraints, but still offer the opportunity for mental model comparison and updating (Sweller & Cooper, 1985; Sweller, Van Merrienboer, & Paas, 1998). Worked examples provide a fully worked-out problem solution to study and explain and can be helpful for developing early foundations for understanding (Atkinson, Derry, Renkl, & Wortham, 2000). Studying with worked examples is generally more effective for novice learning and transfer than practicing conventional problem-solving without assistance (Adams et al., 2014; Nievelstein, Van Gog, Van Dijck, & Boshuizen, 2013; Renkl & Atkinson, 2010; Van Gog, Kester, & Paas, 2011). Further, Hattie (2009) synthesized findings on conventional worked examples and reported an average effect size of d = 0.57, indicating a medium effect.

One hypothesis for the effectiveness of worked examples is that they reduce extraneous processing and cognitive load. Extraneous load relates to the way that information is presented to learners and can be influenced by instruction and the design of learning materials (Sweller et al., 1998). For example, extraneous load may be higher in situations where little guidance is provided such as during conventional problem-solving or when generating sketches. Germane cognitive load refers to the effort required to process the information necessary for learning and construct schemas and mental models. The challenge when designing instruction is to promote generative processing and germane load, while limiting extraneous processing (Adams et al., 2014). Worked examples help manage essential processing and decrease extraneous processing by focusing students' attention and resources on understanding problem solution steps as opposed to irrelevant problem components or ineffective strategies (Zhu & Simon, 1987). While worked examples can free up cognitive resources, they are a relatively passive form of instruction. Freeing up cognitive resources by showing students the solution procedure should allow for greater investment in constructing or updating mental models, but not all students use this freed capacity for generative processing (Renkl & Atkinson, 2010). Thus, worked examples are often coupled with self-explanation prompts in order to facilitate deeper processing and make knowledge explicit (Chi, 2000).

Presenting students with *erroneous examples* where they must find, explain, and fix errors can also encourage deeper processing (Siegler, 2002). Erroneous examples (EEs) present the same steps as a traditional worked example, however, one or more of the steps is incorrect. Generative processing is encouraged by requiring students to locate and explain the errors, then make appropriate corrections. EEs foster students' ability to recognize incorrect strategies, direct attention to common errors, and help students remember and avoid those errors in the future. Recent work suggests explaining a mix of correct and erroneous worked examples is more effective for learning (with medium to large effect sizes, d = 0.50; η_p^2 = .06–.15) than explaining only correct worked examples (e.g., Durkin & Rittle-Johnson, 2012; Große & Renkl, 2007; Huang, Liu, & Shiu, 2008). In fact, Ohlsson (1996) argued that for novices to begin to learn how to solve problems, it is necessary that they first learn how to detect an error, understand what caused it, and explain what is needed to correct it.

Research on EEs has primarily been conducted on mathematical problem-solving. One reason for this is the robust literature characterizing common student errors in early math topics (e.g., Baroody & Ginsburg, 1983; Booth & Koedinger, 2008) which has allowed researchers to generate erroneous examples targeting those misconceptions. Identifying and explaining common errors can help students to recognize and accept when they have chosen incorrect procedures, leading to improved procedural knowledge over practice or correct worked examples alone (Siegler, 2002). The process of explaining errors can draw attention to features that make a procedure inappropriate, reinforce that certain procedures are wrong, and isolate what makes a strategy inappropriate. This process can help replace faulty conceptual knowledge with correct knowledge, which is key to developing expertise (Chi, 2000). Research has also shown efficiency benefits (in terms of time on task and reported mental effort) for erroneous examples compared to conventional problem-solving (e.g., McLaren, Van Gog, Ganoe, Karabinos, & Yaron, 2016; Van Gerven, Paas, Merriënboer, & Schmidt, 2002). For instance, McLaren et al. (2016) demonstrated equal learning outcomes were attained between a tutored problem-solving condition and an erroneous examples condition, while less study time and effort was spent on the intervention problems in the EE condition.

Although laboratory studies support the idea that errorful learning is beneficial (e.g., Kapur, 2014), error avoidance seems to be a driving rule in most American classrooms (Metcalfe, 2017). Stevenson and Stigler (1992) demonstrated differences in how errors are integrated into instruction in eighth grade mathematics lessons across several countries including the United States and Japan. In the United States, a specific set of correct procedures were taught, errors were avoided or ignored, and only correct responses from students were acknowledged or praised. In Japan, where students far outperform the United States in math scores, students first tried to solve problems on their own, then participated in a teacher-led discussion targeting students' problem-solving errors, rather than only recognizing correct procedures. The authors argued that the Japanese method may be more beneficial than the United States method because learning about what is wrong can support understanding about why correct procedures are appropriate and may also foster an environment where students view errors as learning experiences rather than failures.

While there is substantial research on erroneous examples in math learning, little research has focused on the effect of errors or erroneous examples within diagrams. To our knowledge only one study by Wernecke, Schütte, Schwanewedel, and Harms (2018) has investigated the impact of erroneous diagrams on learning. In this study, students were presented with an energy flow diagram that contained no error or contained an error based on a common misconception (that plants get some of their energy from the soil rather than solely from photosynthesis). The error was either circled or students were expected to find it on their own. Students who identified and explained the conceptual error in the diagram learned more about energy flow than students in the correct diagram condition. Therefore, one could hypothesize that students may benefit from identifying and explaining errors in other STEM diagrams, especially when those errors are conceptual in nature. However, many diagrams in STEM, such as 3D geologic block diagrams, are designed to convey spatial information and students have been shown to hold specific spatial misconceptions that result in predictable reasoning errors (Gagnier et al., 2017). An open question is whether identifying and explaining spatial errors can also support learning and comprehension of these kinds of diagrams.

5 | CURRENT STUDY

The goal of the current study was to extend the research on learning from correct and erroneous worked examples beyond basic and applied mathematics domains to a spatial science domain (i.e., 3D geologic block diagrams). Despite the effectiveness and efficiency of EEs for supporting mathematics problem-solving (Booth, Lange, Koedinger, & Newton, 2013; Durkin & Rittle-Johnson, 2012; Große & Renkl, 2007; Siegler, 2002), there has been little research investigating their effectiveness for supporting science learning. Further, there has been no research using EEs in the context of 3D diagram learning or for developing the understanding of spatial information in diagrams. While problem-solving in mathematics typically involves well-defined, multi-step computations, problem-solving in geoscience is often less algorithmic and instead involves making predictions of subsurface structures from surface information, recognizing spatial patterns, and translating between representations (Riggs, Lieder, & Balliet, 2009). Because of this critical difference across the domains, it is an open question whether identifying and explaining erroneous worked examples will show the same benefits in the context of 3D geologic block diagrams.

To carry out this investigation, we compared the effectiveness of an Erroneous Examples condition to two different conditions meant to be analogous to a conventional problem-solving condition and a correct worked examples condition. Sketching served as the conventional problem-solving condition as it is a common problem-solving strategy in STEM and has been shown to support science learning and 3D diagram understanding (Cromley et al., 2020). A correct diagram Copying condition served as the correct worked examples condition. In geoscience, instructors often have students copy block diagram cross-sections into their notes or blank diagram structures. Specifically, students see the initial block, then the cross-section, then see a correct drawing of that cross-section that they must process and copy. This practice is analogous to studying correct worked examples where students see an initial problem, the intermediate stages and final solution, but do not generate that solution on their own.

While generating sketches is both cognitively and practically demanding, it can support learning because it requires deep processing and can facilitate error detection and mental model updating (Gagnier et al., 2017). Erroneous worked examples facilitate mental model updating by isolating what makes certain strategies inappropriate and have been shown to support learning in both mathematics (Durkin & Rittle-Johnson, 2012; Große & Renkl, 2007; Huang et al., 2008) and medical diagnosis (Kopp, Stark, & Fischer, 2008). However, prior work has shown that erroneous examples can also be a more efficient strategy compared to conventional problem-solving strategies (McLaren et al., 2016). Studying correct examples, although imposing less cognitive load (Sweller & Cooper, 1985; Sweller et al., 1998), does not always promote generative processing (Renkl & Atkinson, 2010) and in the case of block diagrams has actually not been an effective strategy for supporting learning (Gagnier et al., 2017).

Based on these ideas and prior findings in the literature we made several predictions for the present study. First, it was predicted that learners would demonstrate a greater pre-to-posttest gain in the sketching activity condition than in the copying condition (Hypothesis 1), replicating findings from Gagnier et al. (2017). Second, it was predicted that learners would also demonstrate a greater a pre-to-posttest gain in the EE activity condition than the copying condition (Hypothesis 2). Concerning potential differences between the three conditions, it was less clear what to expect because both sketching and erroneous examples provide opportunity for spatial feedback, error detection, and mental model updating. Thus, although we did not necessarily think the improvements seen in these two conditions would differ from each other, we did predict that together they would demonstrate greater improvement than copying (Hypothesis 3). However, prior research has shown efficiency benefits (in terms of time on task) for erroneous examples compared to conventional problem-solving (e.g., McLaren et al., 2016; Van Gerven et al., 2002), therefore we did expect erroneous example study to be more efficient (though not necessarily more effective) for learning than sketching (Hypothesis 4).

Finally, because it has been argued that sketching and erroneous examples are beneficial for learning because they help students to identify errors and misconceptions and update their mental models, we planned to conduct an exploratory error analysis. Specifically, we wanted to see if the number and type of corrections students made in the erroneous examples and sketching conditions predicted learning and if different error detection behaviors would be seen across the two conditions.

6 | METHOD

6.1 | Participants

A total of 165 students (122 female; M = 20.40 years, SD = 2.56) from a University in the United States participated in exchange for course credit. The average time on task across the learning intervention conditions was 32 min and 6 s (SD = 3.5 min). Data from five participants was excluded from analysis because their time on task was less than 20 min (more than two SD below the mean) and three were excluded from analysis because their time on task was greater than 41 min (more than two SD above the mean). Additionally, five participants with missing data were excluded, resulting in a final sample of 152 participants. Following a pretest-intervention-posttest design with an immediate posttest, participants were randomly assigned to the *copy* (n = 52), *sketch* (n = 49), or *erroneous examples* (EE) (n = 51) condition. A power analysis conducted using G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007) for a between-subjects repeated-measures ANOVA with power set at 0.80 and α = 0.05 indicated that this sample size would afford sufficient sensitivity to detect medium-size effects ($f \geq 0.25$) as found in prior sketching and erroneous worked examples research. All research was approved by the university's IRB and was conducted in accordance with APA standards for ethical treatment of subjects.

6.2 | Materials

6.2.1 | Geology tutorial

To introduce geology and the importance of 3D diagrams for solving geologic problems, participants saw a self-paced, five-slide Microsoft PowerPoint tutorial explaining geologic block diagrams. The tutorial explained that geologists study changes that occur in rock structures over time, and that block diagrams illustrate rocks that cannot be seen from the Earth's surface. Participants saw example block diagrams with lines indicating where they could be cut to create cross-sections and were told their task would be to think about cross-sections, and what the inside of block diagrams would look like after being cut.

6.2.2 | Geologic block cross-sectioning task

Participants completed a computerized version of the geologic block cross-sectioning task (GBCT; Ormand et al., 2014); a spatial task designed to measure penetrative thinking ability, or the ability to visualize the interior of a 3D structure based on its exterior. In this task, participants are asked to visualize the inside of a geologic block containing multiple layers by mentally "slicing" it along a specified plane (see Appendix A). There are two matched forms, each with 16 multiple-choice items and a time limit of 8 min. Scores were computed as the total correct out of 16. Participants were randomly assigned one version as the pretest and the other as the posttest; order was counter-balanced, and Cronbach's Alpha was acceptable (Form A = .78; Form B = .74).

6.2.3 | Stimuli

The stimuli matched those used by Gagnier et al. (2017), which consisted of color photographs and line drawings of four types of geologic block models and slices through those models. The four models represent common geologic structures that introductory geology students learn about: dipping layers, horizontal layers cut by a reverse fault, and gently and steeply plunging folds. All stimuli were presented in PowerPoint on a 24-in flat-screen Dell monitor with participants seated approximately 50 cm away.

6.2.4 | Learning intervention

Participants in all three conditions completed a self-paced learning intervention during which they viewed a PowerPoint presentation showing vertical cuts (i.e., cuts that were straight down into the model) into four Play-Doh models of geologic structures (Figure 1). Each structure was cut three times resulting in a total of 12 possible geologic block items. The copy and sketching interventions matched those used in Gagnier et al. (2017); the EE intervention was developed for the present study.

In the sketching condition students viewed a model (Appendix B) and generated a sketch of what they thought the model would look like following the indicated cut. After each sketch, students wrote explanations of how they used the layers visible on the front, side and top of the block to make their prediction (Appendix C). Next, they viewed the correct answer and a diagram of the sliced block, compared their sketch to the correct answer, and indicated what differed between their sketch and the correct answer. The correct model and diagram were spatially aligned with the original model to promote comparison. This sequence was repeated for Cuts 2 and 3 into the model. After completing the sketch for Cut 3 and comparing it to the correct answer, a new model appeared, and this sequence was repeated for the next three models for a total of 12 possible sketches or until 10 min remained in the hour for completing the posttest and debriefing.

In the copy condition, participants viewed the same images in the same order, but were also shown the block after it was cut and instructed to copy the structure into the blank block template provided in the packet. Participants were told to sketch each layer individually on all three visible sides of the block, and were asked to write explanations of how they used the layers visible on the front, side and top of the block to make their sketches (these sketches were *not predictive* because students were able to create while seeing the actual cross-section). After making their copy, they saw the correct sketch and marked any differences between their drawing and the correct drawing. Hence, this group generated diagrams of the same structures as the sketching group, were prompted to explain their process, and then were prompted to correct any errors they saw in their sketches but did not engage in visualization or prediction.

The stimuli for the EE condition matched the other conditions, with exception of the task booklet. The task booklets contained block diagrams that were completed with sketches containing various errors. Participants were told they were made by previous students and represent predictions of what they thought the face would look like after the cut. Participants identified and marked any errors they could find and wrote explanations how the students' sketches were incorrect on all three sides of the block. After identifying and explaining errors, participants saw the correct model and diagram and marked any additional errors they noticed. To create the erroneous sketch stimuli, student sketches from Gagnier et al. (2017) were analyzed. The most common errors made by previous participants, in order of frequency, included errors in the shape and thickness of the layers, incorrect inclusion of an additional layer, exclusion of a layer, errors in connecting the layers on each side of the block, and errors in overall pattern alignment. Each erroneous sketch included approximately three common errors. Across all 12 sketches there were 38 possible errors for students to identify.

6.2.5 | Demographic form

Participants were asked to report their age, gender, education level, bilingual status, and the number and name of college science courses taken (Table 1).

6.3 | Procedure

Upon arrival, participants were assigned an identification number and completed a consent form. Next, they were told they would be introduced to geology and were instructed to read through the tutorial because the information would be useful for the study. After the tutorial, participants completed the pre-GBCT and then completed the self-paced learning intervention. In all conditions, participants saw the stimuli in the same order. After completing the learning intervention, or being prompted to stop because they reached the final 10 min of the session, participants completed the post-GBCT. Lastly, they completed the demographic questionnaire, were debriefed, and thanked for their participation. The study took 1 hour.

6.3.1 | Error coding

To examine the effect of explaining relevant spatial errors on learning how to interpret 3D geologic block diagrams, each sketch was coded for the presence of common errors. This coding focused on errors students made *and* identified or explained in their sketches. In the copy and sketching conditions, each sketch was coded for the presence or absence of six possible error types: (a) whether the predicted shapes of the layers were correct, (b) whether the predicted thicknesses of the layers were correct, (c) whether any layers were excluded, (d) whether any extra layers were present, (e) whether the layers drawn on all three sides were connected, and (f) whether the overall alignment of the block pattern was correct. Because each sketch was coded for six possible errors, a maximum of 72 errors was possible, but no students made or explained this many; the maximum errors seen in the sketching condition was 58 and in the copy condition was 49. Example sketches can be seen in Figure 2.

Reliability for sketch coding was established by having a second coder score 30% of the sketches (approximately 450 sketches). For the copy and sketch conditions Krippendorff's alpha reliability coefficients were high (Sketch 1: α = .89; Sketch 2: α = .87; Sketch 3: α = .93). In the EE condition, because students did not generate drawings, their packets were coded for whether they identified and explained each error present in each sketch. Krippendorff's alpha reliability coefficients for sketches in the EE condition were satisfactory (Sketch 1: α = .79; Sketch 2: α = .82; Sketch 3: α = 79). Disagreements were settled by a third coder. A breakdown of error type by sketch can be seen in Table 2. Although 38 errors were possible in the EE condition, the most anyone identified was 32 (see Table 3 for mean errors identified).

TABLE 1 Descriptive statistics as a function of the intervention activity condition

	Copy (n = 52)	Sketch (n = 49)	Erroneous example (n = 51)
Time (s)	1919.96 (204.51)	2077.43 (217.10)	1831 (353.43)
Sketches completed	7.60 (3.32)	7.94 (3.03)	11.25 (1.60)
Age	21.33 (3.36)	20.14 (1.37)	20 (1.99)
% females	0.75 (0.44)	0.82 (0.39)	0.73 (0.45)
Educational level	1.75 (1.33)	1.61 (1.30)	1.14 (1.06)
# science courses	3.23 (3.37)	2.84 (3.10)	2.41 (2.39)

Note: SD are presented in parentheses following the mean. Educational level: 1, freshman; 2, sophomore; 3, junior; and 4, senior.

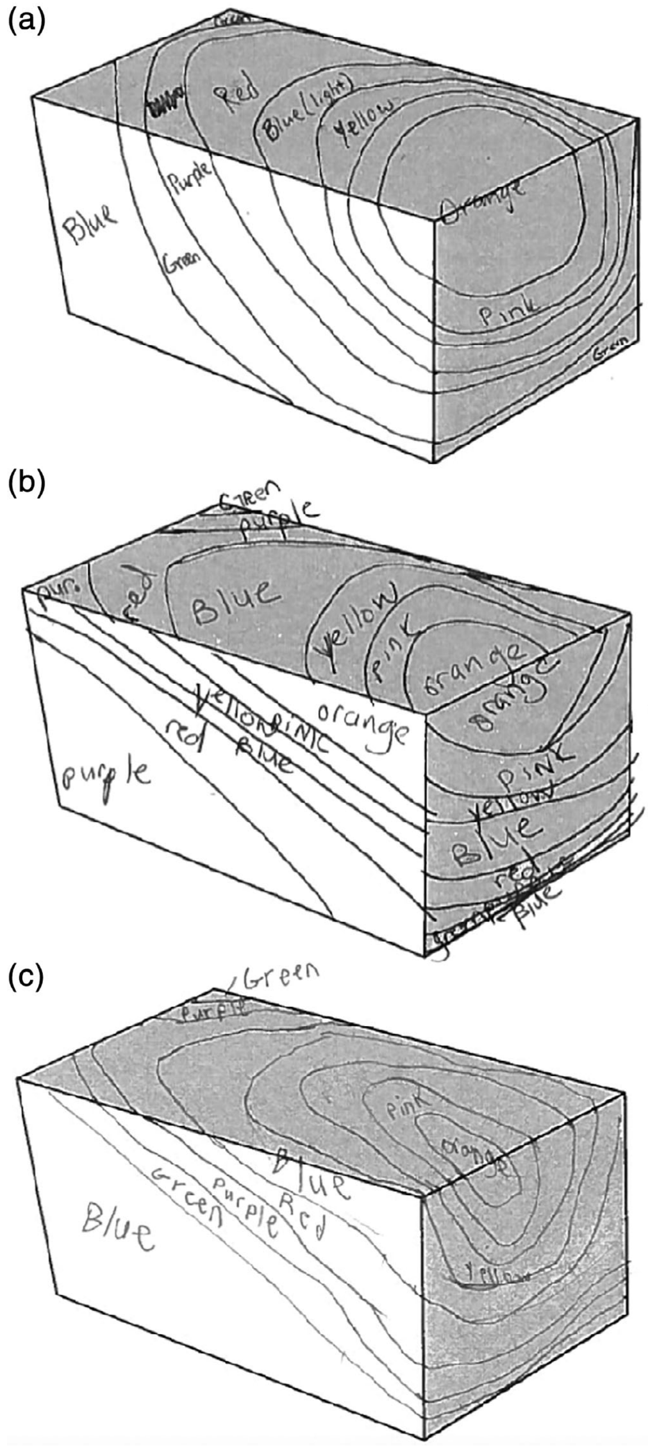


FIGURE 2 Examples of diagrams in the sketching condition, illustrating the various errors. (a) Shape and thickness errors on the front face; (b) Connection and missing layers errors on the front face; (c) Alignment and extra layer errors on the perpendicular face

7 | RESULTS

7.1 | Performance on geologic block cross-sectioning test

To examine whether identifying and explaining errors in erroneous examples improves penetrative thinking skill, pre- to posttest scores were compared across learning intervention conditions. A one-way Analysis of Variance (ANOVA) found no difference in pretest score across activity intervention conditions indicating that conditions were matched, $F < 1$. Next, a 2 (Test type: pre-, posttest) X 3 (Condition: copy, sketch, EE) repeated-measures ANOVA was conducted and revealed a main effect for test type indicating performance on the posttest was greater than performance on the pretest, $F(1,149) = 10.84$, $p < .001$, $\eta_p^2 = .07$. There was no effect for activity condition, $F < 1$, and there was a marginal interaction between test type and activity condition, $F(2,149) = 2.47$, $p = .08$, $\eta_p^2 = .03$. See Table 4 for unadjusted means.

To directly test hypotheses 1 through 3, a series of contrast analyses were conducted comparing pre-to-posttest gains across each intervention condition. Contrary to our prediction and the results obtained by Gagnier et al. (2017), the pre-post gain in the sketching condition did not significantly differ from that of the copying condition, $t(149) = 1.52$, $p = .13$ (Hypothesis 1). However, pre-post gain in the EE condition was greater than that of the copying condition, $t(149) = 2.09$, $p < .04$, $d = .34$ (Hypothesis 2). Because it was hypothesized that both sketching and EEs would be beneficial, an additional planned contrast was conducted comparing pre- to posttest gain in the copying condition to the total gain seen across both the sketching and EE conditions (−2, 1, 1), which revealed that the average improvement seen across the sketching and EE conditions was significantly greater than that of the copying condition, $t(149) = 2.09$, $p < .04$, $d = .34$ (Hypothesis 3), but the improvement in the sketching condition did not differ from the improvement in the EE condition, $t < 1$.

7.2 | Number of sketches completed

The mean number of sketches or diagrams completed was 7.60 ($SD = 3.32$) in the copy condition, 7.94 ($SD = 3.03$) in the sketch condition, and 11.25 ($SD = 1.60$) in the EE condition. A one-way ANOVA revealed that this difference was significant, $F(2,149) = 27.47$, $p < .001$, $\eta_p^2 = .27$ (Table 1). Post hoc Tukey Honestly Significant Difference tests (HSD) indicated that participants in the sketch and copy conditions did not differ in the total number of sketches completed (*ns*) while participants in the EE condition completed significantly more sketches than participants in both the sketch and copy conditions (*ps* < .001). Thus, participants in the EE condition completed on average 3.31 more sketches than students in the sketching condition and 3.66 more than students in the copy condition.

One interpretation of these results is that participants in the EE condition improved on diagram understanding because they got more

TABLE 3 Error correction frequency as a function of learning intervention condition

	Copy	Sketch	Erroneous example
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Total errors made/possible	15.46		(10.00)
22.86 (11.37)	35.05		(6.21)[a]
3D structure errors made/possible	4.71 (5.44)	10.06 (7.63)	15.73 (2.71)[a]
2D surface errors made/possible	10.73 (5.32)	12.80 (4.54)	19.39 (3.48)[a]
Total errors corrected	4.85 (4.31)	9.96 (6.00)	20.67 (5.04)
Proportion corrections	0.35 (0.27)	0.46 (0.19)	0.60 (0.13)
3D structure error corrections	1.08 (1.79)	3.82 (3.33)	12.80 (2.79)
2D surface error corrections	3.77 (3.12)	6.14 (3.12)	7.86 (3.19)
Proportion of 3D structure errors corrected out of 3D structure errors made/possible	0.22 (0.33)	0.44 (0.27)	0.82 (0.12)
Proportion of 2D surface errors corrected out of 2D surface errors made/possible	0.38 (0.30)	0.50 (0.22)	0.42 (0.17)

[a]In the erroneous examples condition, students did not make errors, but rather were given diagrams with errors already present in them. Thus, the values for total errors made, 3D structure errors made, and 2D surface errors made in the erroneous examples condition reflect how many errors were possible to correct or explain.

TABLE 2 Error type by sketch number in the erroneous example condition

	Item #											
Error type	1	2	3	4	5	6	7	8	9	10	11	12
Missing layer	✓	✓						✓		✓	✓	✓
Extra layer			✓	✓	✓	✓	✓	✓	✓		✓	
Thickness		✓	✓				✓		✓	✓	✓	✓
Shape	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓
Connection	✓			✓	✓				✓	✓	✓	
Alignment												✓

TABLE 4 Unadjusted pre and post geologic block cross-sectioning test means and *SD* as a function of intervention activity condition

	Copy	Sketch	Erroneous example
Pre	5.79 (3.30)	5.59 (3.14)	5.08 (3.06)
Post	5.85 (3.73)	6.45 (3.70)	6.22 (3.72)

practice with geologic block diagrams than participants in the sketching or copy conditions. However, two findings suggest this is not the case. First, replicating Gagnier et al. (2017), there was no relationship between number of sketches or diagrams completed and pre-to-posttest gain ($r = -.05$, *ns*). Second, a 2 (test type) X 3 (learning intervention) ANCOVA controlling for the number of sketches completed revealed a significant main effect for test type, $F(1,148) = 6.65$, $p = .01$, $\eta_p^2 = .04$, and still no main effect for condition, $F < 1$. Importantly, there was no main effect of the covariate, $F(1,148) = 1.59$, $p = .21$, and no interaction between the covariate and the within-subjects factor (pretest and posttest; $F(1,148) = 2.84$, $p = .09$) or between the covariate and the between-subjects factor (intervention condition; $F(2,146) = 1.24$, $p = .29$) indicating that the assumption of homogeneity of regression slopes was not violated. However, now a significant interaction between test type and intervention condition was obtained, $F(2,148) = 3.76$, $p < .03$, $\eta_p^2 = .05$, suggesting that gain from pre- to posttest varied as a function of activity intervention even when controlling for the number of sketches completed (Figure 3).

To further assess whether the number of sketches a student completed impacted their pre-to-post GBCT gain, a mediation model between activity condition, number of sketches completed, and pre-to-posttest gain was tested. We used the PROCESS macro (version 3.3; Hayes, 2013) for SPSS to test the effect of a mediating variable on the relationship between a multicategory (i.e., noncontinuous) independent variable and a continuous dependent variable (Hayes & Preacher, 2014). X1 contrasted the sketching and copying conditions and X2 contrasted the sketching and EE conditions. As shown in Figure 4, the unstandardized regression coefficients outside the parentheses reflect the effect of intervention condition on pre-post gain, and the coefficients inside the parentheses reflect when number of sketches is included in the model. To test for significant indirect effects, we used bootstrapping with 5,000 resamples to obtain bias-corrected 95% confidence intervals. If zero is outside the confidence intervals, the indirect effect is consequently not zero and can thus be interpreted as evidence of mediation (Preacher & Hayes, 2008). However, the confidence intervals resulting from this analysis for both X1 (95% CI [−.13, .31]) and X2 (95% CI [−1.03, .09]) did contain zero, suggesting that number of sketches did not mediate the relation between intervention condition and GBCT pre-post gain.

Overall, these results indicate that although participants in the EE condition completed more sketches than participants in the copy or sketch conditions, there is no evidence that simply completing more sketches led to the observed learning gains. As shown in Figure 5, it is unlikely that the observed effects were driven by outliers or extreme scores.

7.3 | Efficiency benefits

Differences in overall time on task as a function of activity intervention condition was also investigated. Although we attempted to equate time on task across conditions by setting a required stoppage time, a one-way ANOVA revealed a significant difference across conditions, $F(2,149) = 10.87$, $p < .001$, $\eta_p^2 = .13$ (Table 1). Follow-up Tukey HSD tests demonstrated that participants in the sketching condition spent more time on task than participants in the copy condition ($p < .01$) and the EE condition ($p < .001$) (Hypothesis 4). The difference in time on task between the erroneous example and copying conditions did not reach significance. This also suggests that the improvement seen in the EE condition was not due to merely more exposure to the intervention.

7.4 | Exploratory error analysis

The fact that pre-post gains were seen in both the sketching and EE conditions suggests that perhaps the act of identifying and explaining

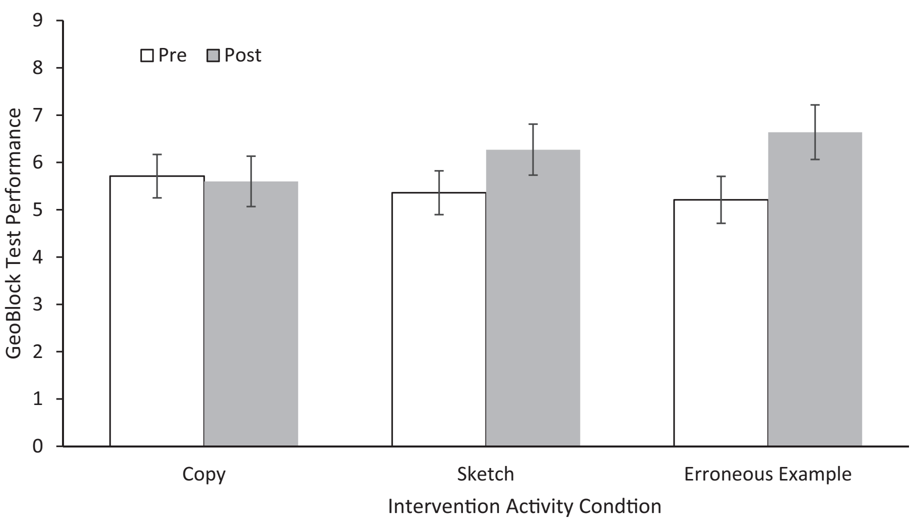


FIGURE 3 Pre- and posttest GBCT performance (adjusted means from ANCOVA) as a function of intervention activity. Error bars represent the *SE*. GBCT, geologic block cross-sectioning task

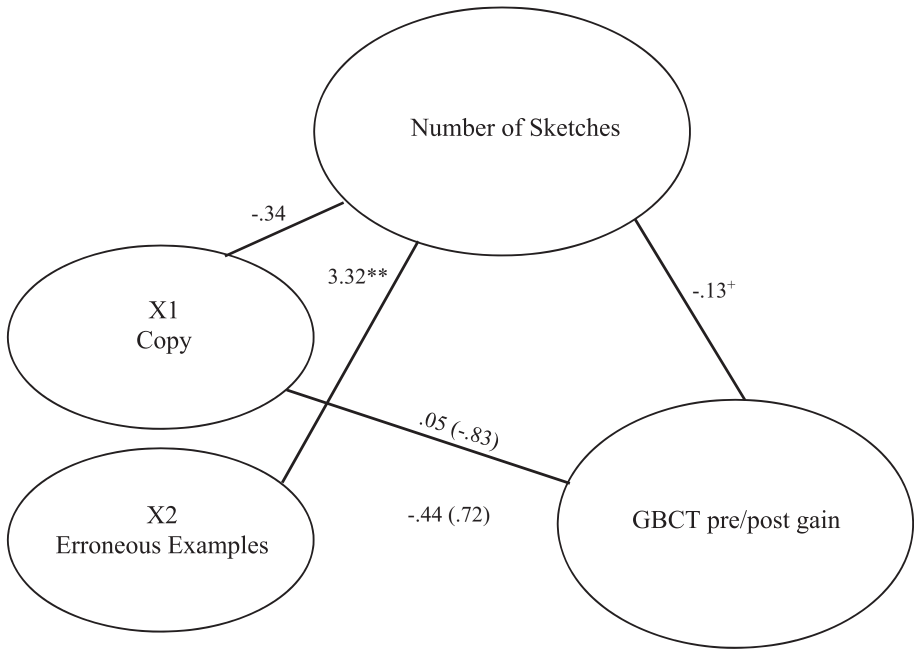


FIGURE 4 Mediation model including intervention condition, number of sketches completed, and pre-to-post GBCT gain. GBCT, geologic block cross-sectioning task

errors, either self-generated or provided, and updating one's own mental representation to reflect those corrections is a critical piece of what makes these learning activities effective. To more directly evaluate this assertion, an exploratory error analysis was conducted. Because there were differences in the number of errors participants made across conditions (see Table 3) and thus students in some conditions had more opportunity for error correction, a proportion score was computed for each participant that represented their total number of error corrections as a function of how many errors they made (or, in the EE condition, the number of errors they corrected out of the total number of errors present). A one-way ANOVA examining proportion of errors corrected out of errors made as a function of activity condition revealed an overall effect, $F(2,149) = 19.89$, $p < .001$, $\eta_p^2 = .21$. Follow-up Tukey HSD tests revealed that

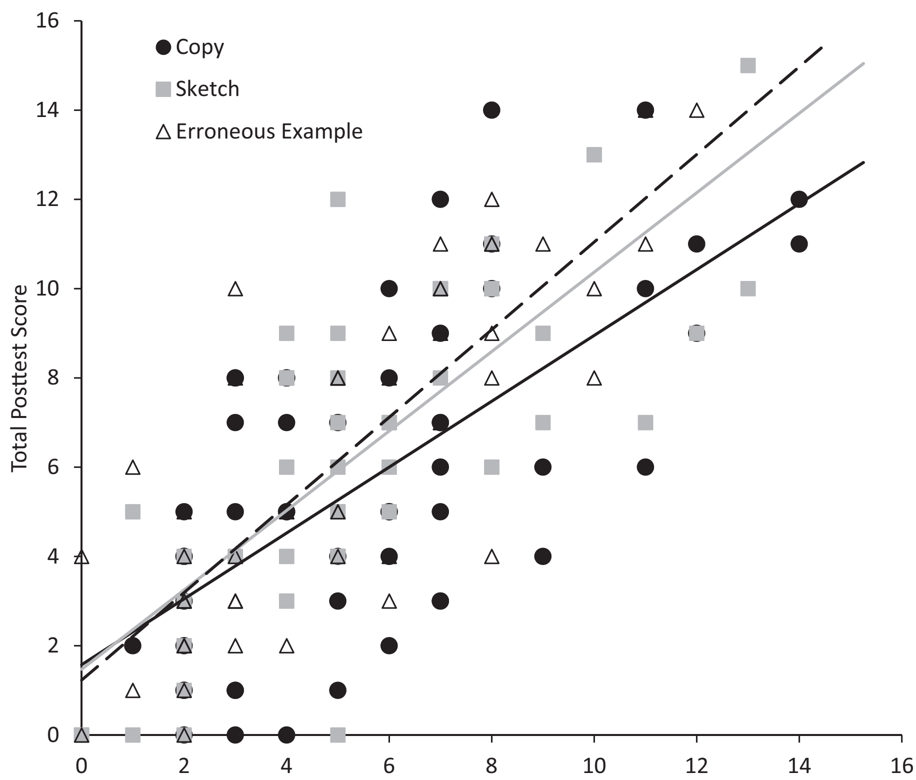


FIGURE 5 The relationship between pre- and posttest GBCT scores as a function of activity condition (solid black line, copy; sold gray line, sketch; dotted line, erroneous example). GBCT, geologic block cross-sectioning task

participants in the EE condition (M = .60, SD = .13) explained a greater proportion of the errors present than participants in the sketching condition (M = .46, SD = .19; p < .002) and the copying condition (M = .35, SD = .27; p < .001), and participants in the sketching condition explained and corrected a greater proportion of the errors they made than participants in the copying condition (p < .03).

Differences in the type of errors students explained and corrected as a function of activity condition was also assessed. To do so, the six error categories were combined into two major categories: errors related to the overall 3D *structure* of the block and its layers and errors related to 2D *surface features* of the block. Structure errors were defined as those that required students to think about multiple sides simultaneously or the entire 3D structure of the block. For example, errors in alignment or connection were 3D in nature because they impacted the entire block as opposed to only a single side. Similarly, errors in the number of layers represented in the block also impacted the structure of the entire block. Thus, these four categories of errors were combined to represent Structure errors. Surface errors were defined as those that did not require students to think about multiple sides simultaneously or the 3D structure, but rather were 2D in nature and occurred on a single surface. For example, errors in the shape or thickness of layers could occur on only a single side and could be unrelated to the shapes or thicknesses seen on another side. In the EE condition across all 12 sketches, there were 17 possible 2D surface errors and 21 possible 3D structure errors. Mean number of 2D surface and 3D structure errors made by participants in the Sketching and Copying conditions can be seen in Table 3 as well as mean number of errors possible in the EE condition.

Two proportion scores were created for each participant, one represented the number of 2D surface corrections as a function of total 2D errors made/possible and the other represented the number of 3D structure corrections as a function of total 3D errors made/possible. A 2 (Error type) X 3 (Learning intervention) ANOVA examining proportion of 2D and 3D errors corrected out of errors made as a function of activity condition revealed a main effect for error type, $F(1,149) = 7.53$, $p < .01$, $\eta_p^2 = .05$, which indicated that students corrected a greater proportion of the 3D structure errors they made (M = 0.49, SD = 0.35) than the 2D surface errors (M = 0.43, SD = 0.24). Like the overall error analysis conducted initially, there was a main effect for intervention condition, $F(2,149) = 30.20$, $p < .001$, $\eta_p^2 = .29$. These main effects were qualified by a significant error type by intervention condition interaction, $F(2,149) = 64.36$, $p < .001$, $\eta_p^2 = .46$. Follow-up pairwise comparisons using Bonferroni correction revealed that participants explained and corrected a greater proportion of the 2D surface errors they made than 3D structure errors in the copy condition, $t(51) = 3.46$, $p < .001$, $d = .48$. In the sketching condition, participants explained and corrected an equal proportion of the 2D and 3D errors they made, $t(48) = 1.55$, $p = .13$ (Table 3). On the other hand, in the EE condition participants identified and explained a greater proportion of structure errors present in the erroneous examples than surface errors, $t(50) = 17.78$, $p < .001$, $d = 2.49$.

8 | DISCUSSION

Previous research demonstrates that sketching can improve penetrative thinking and reasoning about geologic block diagrams (Gagnier et al., 2017), however sketching can be challenging for many students because it is cognitively demanding and time-consuming (Garnier et al., 2017; Jaeger et al., 2018). Thus, while sketching is a common practice in many STEM domains, it may not always be an appropriate instructional tool in the classroom because of the time and effort constraints it imposes. The present study sought to explore whether explaining and correcting errors in erroneous examples would show similar or greater learning benefits to a sketching activity, but also be a more efficient instructional tool. To our knowledge, this is only the second study to extend the erroneous worked examples paradigm to STEM diagram learning, and the first to focus on 3D spatial information in diagrams.

Results indicated that explaining erroneous examples led to a greater pre-to-posttest gain in penetrative thinking than copying correct diagrams. This result supports the idea that identifying and explaining erroneous examples can support learning, but importantly, extends these results beyond mathematics to 3D geologic diagram learning. Interestingly though, while the improvement seen in the erroneous examples condition was significantly greater than the improvement in the copying condition, the improvement in the sketching condition did not significantly differ from the improvement in the copy condition. This result does not replicate the benefits for sketching found by Gagnier et al. (2017) and suggests that erroneous examples may be a more effective pedagogical tool for supporting geologic block diagram understanding than sketching. However, it is important to note that when compared to each other, the pre-to-posttest gains in the sketching and erroneous examples conditions did not differ.

The results of the present study also demonstrated that participants in the EE group completed more items than participants in the copying and sketching groups, but also spent significantly less time on task. These results align with prior research showing worked examples to be more efficient than traditional problem-solving practice (e.g., McLaren et al., 2016; Van Gerven et al., 2002) and suggests that in the context of real classrooms where time is often very limited, erroneous examples may be a more useful tool for instructors.

Results from the exploratory error analyses are useful in trying to explain and understand the pattern of results obtained for pre- and posttest performance on the Geological Block Cross-Section Test. Results indicated that participants in the copying condition corrected or explained very few of the errors they made in their drawings. Further, of the errors they did correct, most of them were 2D surface errors as opposed to larger 3D structural errors. These error results, and the fact that students in the copying condition showed no pre-to-posttest improvement, suggest that copying correct examples, and correcting errors made in one's copies may not be fostering the same level of attention to error identification and mental model updating as the sketching and erroneous examples conditions. Another possibility for the lack of gains seen in the copying condition could stem from the explanations students were prompted to make prior to the error correction phase. In the copying condition, students were asked to write explanations of how they used the information on the visible sides of the block to make their sketches. Because students were essentially copying what they saw into the blank block template, the self-explanations regarding this process may not have been effective for deep processing or mental model building.

One argument for why both sketching and erroneous examples are effective instructional tools is that they allow for error identification and correction and mental model updating. The fact that the pre-to-post gains seen in the erroneous examples and sketching groups did not differ from each other does seem to lend some support for this argument. However, results from the exploratory error analysis indicate that the explanation is likely more complicated. Specifically, the error analyses indicated that participants in the EE group corrected and explained a greater proportion of errors than participants in the sketching group. Further, of the errors that were explained and corrected, the erroneous examples group explained more 3D structural errors than 2D surface errors while participants in the sketching condition focused equally on explaining both types of errors.

Because the error analysis was exploratory, it is difficult to make concrete inferences about how attention to errors impacted learning about the geologic block diagrams. Improvements in geologic block diagram understanding did not differ between the sketching and EE conditions despite differences in the number and type of errors corrected. Together, these results do suggest that the benefit of sketching does not merely lie in the opportunity for error identification and correction, but that there is something unique in the act of generating a predictive sketch that facilitates mental model creation. For example, perhaps it is the case that generating a sketch requires the learner to think about the 3D structure of the block, and thus, even if they do not attend to 3D errors in a correction phase, learners already have a 3D representation in their mental model. On the other hand, because it is not necessary to develop a 3D mental model in the erroneous examples condition because a representation is already provided, attention to 3D errors may be critical for supporting students reasoning about the 3D structure of the block. Unfortunately, the present data cannot offer concrete evidence to support these hypotheses and future research is required. An important open question is whether the two pedagogical tools should be combined in instruction. Erroneous examples could support students who might be initially struggling to synthesize meaning from complex diagrams and efficiently address high probability errors, and sketching could support independent mental model development, and unconstrained exploration of a diagram's potential for representing spatial information.

9 | LIMITATIONS AND FUTURE RESEARCH

Several limitations in the present study should be addressed in future research. First, the learning measure used was very near to the materials used in the intervention. This begs the question of whether transfer to other complex geology diagrams or applied penetrative thinking scenarios would be possible. Relatedly, this study investigated a single type of 3D diagram, but 3D diagrams are present across STEM. Future research should extend this paradigm to evaluate whether erroneous examples can support learning of other 3D diagrams such as molecular diagrams in chemistry or cellular structure in biology. Further, future work should more carefully consider the self-explanation component in each condition. Research on worked examples has shown that a critical part of what makes them effective for supporting learning is the self-explanation prompts that are included alongside the worked examples (Renkl & Atkinson, 2010). In the present study, the copying condition was meant to be analogous to a correct worked examples condition. Participants were prompted to explain how they used the information from each visible side of the geologic block diagram to make their sketches, but this prompt was rather shallow and may not have elicited the same kind of reflective thinking and mental model development fostered by other self-explanation prompts.

Because a relationship between STEM and cognitive capacities such as spatial thinking (Uttal & Cohen, 2012; Wai, Lubinski, & Benbow, 2009) and working memory (Danili & Reid, 2004; Gathercole, Pickering, Knight, & Stegmann, 2004) has been demonstrated, future research should investigate whether erroneous examples are especially supportive for students with less of these cognitive resources. Jaeger et al. (2018) argue that students with low-spatial skills may struggle to generate spatial mental models and therefore may also struggle to generate sketches. Thus, erroneous examples could be especially beneficial for low-spatial students because they provide an initial visual representation that can be evaluated rather than requiring the student to generate their own. Similarly, since erroneous examples can reduce cognitive load, they may be useful tools for students with low working memory capacity. Future research should include individual differences measures including spatial thinking skills and working memory capacity, measures of prior knowledge, and measures of cognitive load, to address these questions.

Although the present study found that erroneous examples were more beneficial than copying correct diagrams, it did not show that erroneous examples led to greater gain than sketching. Thus, it is not the case that erroneous examples were clearly a *better* learning strategy than sketching in terms of pre-to-post improvement. Therefore, future research should tease apart when sketching activities versus erroneous examples problems would be a better instructional strategy. Some issues to be considered in future research include differences in the complexity of the to-be-learned material (e.g., Große & Renkl, 2007), the kinds of corrective feedback or direct instruction students receive (e.g., Kapur, 2014), and if there is an optimal amount of time or number of items that should be included for sketching and erroneous examples activities. In the present study, students were made to stop the learning activity when 10 min remained in the session. Many students in the sketching and copying conditions did not complete all 12 items and thus, an open question is whether completing all 12 items would have led to further pre-to-posttest improvement in either of these two groups.

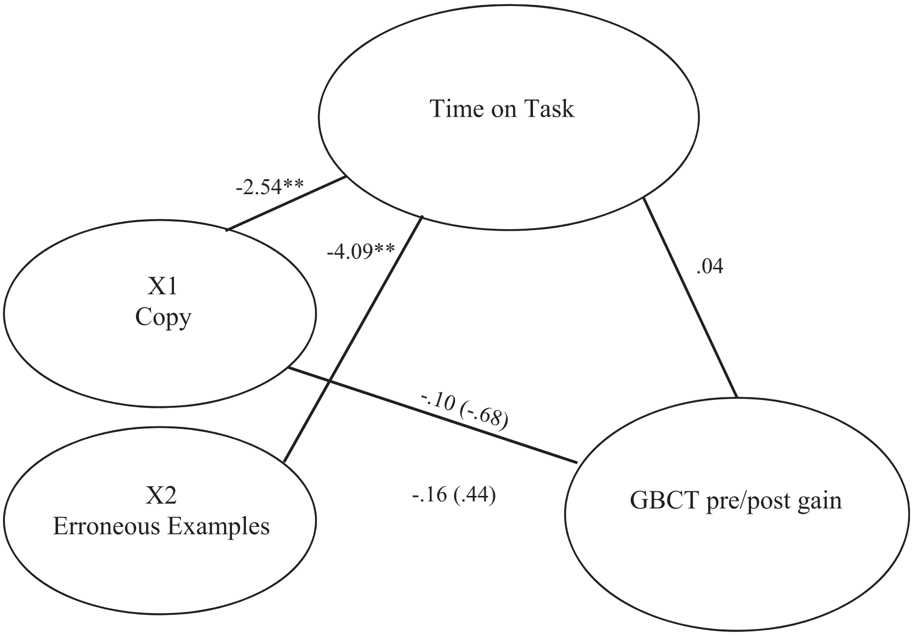


FIGURE 6 Mediation model including intervention condition, time on task, and pre-to-post GBCT gain. GBCT, geologic block cross-sectioning task

Because designing erroneous examples requires including errors most likely to occur, having a robust understanding of what types of errors occur for which diagrams is critical. Unfortunately, visual literacy activities that focus on generating visual representations or interpreting erroneous representations are rarely implemented in the classroom. When teaching previously under-researched topics, student sketching could be employed to characterize errors providing a practical, and ideally theoretical, basis for item construction.

Despite these limitations, this study provides an innovative approach to developing students penetrative thinking skills and 3D diagram comprehension. Considering students' exposure to 3D diagrams across STEM domains and their use for conveying complex processes, it is essential for students to not only develop the ability to understand them, but also to create them and use them as tools for supporting mental model generation and updating. Unfortunately, instruction on diagram interpretation and generation does not exist in many curricula, in part because generating sketches or spatial representations is inefficient and impractical within the time constraints of the classroom. The results of the present study indicate that a more practical classroom activity for supporting 3D diagram comprehension is examining and explaining errors in erroneous worked examples. Future research should continue to investigate the effectiveness and efficiency of erroneous worked examples for fostering diagram comprehension and penetrative thinking skills and importantly, should continue to investigate the mechanisms that make these learning activities beneficial.

ACKNOWLEDGMENTS

This research and the preparation of this article was supported in part by National Science Foundation Grant SBE 10-41707 to the Spatial Intelligence and Learning Center (SILC) and in part by National Science Foundation Grants 1640800 and 1839705. We thank Nahal Heydari, Mia Velazquez, Nora Newcombe, and other members of SILC for feedback throughout the process. We also thank research assistants Shayna Black, Breanna Orlando, and Julie Schilling.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Allison J. Jaeger https://orcid.org/0000-0002-6238-8923

ENDNOTES

[1]Correlations between number of sketches completed and pre-post gain were run for each condition separately and no significant relationships were present (Copy: r = −.12, p = .37; Sketch: r = −.21, p = .15; EE: r = −.05, p = .73).

[2]A mediation model between activity condition, time on task, and pre-to-posttest gain was also tested. Model 1 contrasted the sketching and copying conditions and model 2 contrasted the sketching and EE conditions (Figure 6). The confidence intervals resulting from this analysis for both model 1 (X1 indirect = −.13, SE = .15, 95% CI [−.42, .16]) and model 2 (X2 indirect = −.06, SE = .08, 95% CI [−.25, .09]) did contain zero, suggesting that time on task did not mediate the relation between intervention condition and GBCT pre-post gain.

[3]A reviewer suggested we also follow-up the interaction by reporting whether our groups differed from each other in terms of proportion of 2D and 3D errors corrected. To assess this, we conducted two separate one-way ANOVAs. When looking at proportion of 2D errors corrected as a function of activity condition, there was a significant effect, $F(2,149) = 3.16$, $p = .04$, $\eta_p^2 = .04$. Follow-up Tukey HSD tests showed that there was no difference in the proportion of 2D errors corrected between the copy and EE conditions ($p = .75$), or between the EE and sketch conditions ($p = .20$), but participants corrected a greater proportion of 2D errors in the sketch condition than the copy condition ($p = .04$). When looking at proportion of 3D errors corrected as a function of activity condition, there was a significant effect, $F(2,149) = 72.01$, $p < .001$, $\eta_p^2 = .49$. Follow-up Tukey HSD tests showed that all three conditions significantly differed from each other in terms of proportion of 3D errors corrected, with the EE condition showing the greatest proportion of 3D errors corrected and the

copying condition showing the lowest proportion of 3D errors corrected (all p's < .001).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Allison J. Jaeger https://orcid.org/0000-0002-6238-8923

REFERENCES

Adams, D. M., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., & Van Velsen, M. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior*, *36*, 401–411.

Ainsworth, S., Prain, V., & Tytler, R. (2011). Science education: Drawing to learn in science. *Science*, *333*, 1096–1097.

Alesandrini, K. L. (1981). Pictorial–verbal and analytic–holistic learning strategies in science learning. *Journal of Educational Psychology*, *73*, 358–368.

Alles, M., & Riggs, E. M. (2011). Developing a process model for visual penetrative ability. *Geological Society of America Special Papers*, *474*, 63–80.

Atit, K., Gagnier, K., & Shipley, T. F. (2015). Students' gestures help their penetrative thinking skills. *Journal of Geoscience Education*, *63*, 66–72.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*, 181–214.

Baroody, A., & Ginsburg, H. (1983). The effects of instructions on children's understanding of the equals sign. *The Elementary School Journal*, *84*, 199–212.

Booth, J. L., & Koedinger, K. R. (2008). Key misconceptions in algebraic problem solving. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual cognitive science society* (pp. 571–576). Austin, TX: Cognitive Science Society.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24–34.

Britton, L. A., & Wandersee, J. H. (1997). Cutting up text to make moveable, magnetic diagrams: A way of teaching and assessing biological processes. *The American Biology Teacher*, *59*, 288–291.

Brunyé, T. T., & Taylor, H. A. (2008a). Working memory in developing and applying mental models from spatial descriptions. *Journal of Memory and Language*, *58*, 701–729.

Brunyé, T. T., & Taylor, H. A. (2008b). Extended experience benefits spatial mental model development with route but not survey descriptions. *Acta Psychologica*, *127*, 340–354.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Erlbaum.

Coleman, J. M., McTigue, E. M., & Smolkin, L. B. (2011). Elementary teachers' use of graphical representations in science teaching. *Journal of Science Teacher Education*, *22*, 613–643.

Cooper, M. M., Stieff, M., & DeSutter, D. (2017). Sketching the invisible to predict the visible: From drawing to modeling in chemistry. *Topics in Cognitive Science*, *9*, 902–920.

Cromley, J. G., Bergey, B. W., Fitzhugh, S., Newcombe, N., Wills, T. W., Shipley, T. F., & Tanaka, J. C. (2013). Effects of three diagrams instruction methods on transfer of diagram comprehension skills: The critical role of inference while learning. *Learning and Instruction*, *26*, 45–58. https://doi.org/10.1016/j.learninstruc.2013.01.003

Cromley, J. G., Du, Y. & Dane, A. P. (2020). Drawing-to-learn: Does meta-analysis show differences between technology-based drawing and paper-and-pencil drawing? *Journal of Science Education and Technology*, *29*, 216–229.

Danili, E., & Reid, N. (2004). Some strategies to improve performance in school chemistry, based on two cognitive factors. *Research in Science and Technological Education*, *222*, 203–226.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, *54*, 207–231.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning & Instruction*, *22*, 206–214.

Fan, J. E. (2015). Drawing to learn: How producing graphical representations enhances scientific thinking. *Translational Issues in Psychological Science*, *1*, 170–181.

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

Gagnier, K. M., Atit, K., Ormand, C. J., & Shipley, T. F. (2017). Comprehending 3D diagrams: Sketching to support spatial reasoning. *Topics in Cognitive Science*, *9*, 883–901.

Garnier, B., Chang, M., Ormand, C., Matlen, B., Tikoff, B., & Shipley, T. F. (2017). Promoting sketching in introductory geoscience courses: CogSketch geoscience worksheets. *Topics in Cognitive Science*, *9*, 943–969.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *181*, 1–16.

Gobert, J. D. (2005). The effects of different learning tasks on model-building in plate tectonics: Diagramming versus explaining. *Journal of Geoscience Education*, *53*, 444–455.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning & Instruction*, *17*, 617–634.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York, NY: Guilford Press.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*, 451–470.

Huang, T. H., Liu, Y. C., & Shiu, C. Y. (2008). Construction of an online learning system for decimal numbers through the use of cognitive conflict strategy. *Computers & Education*, *50*, 61–76.

Jaeger, A. J., Velazquez, M. N., Dawdanow, A., & Shipley, T. F. (2018). Sketching and summarizing to reduce memory for seductive details in science text. *Journal of Educational Psychology.*, *110*, 899–916. https://doi.org/10.1037/edu0000254

Johnson, J. K., & Reynolds, S. J. (2005). Concept sketches using student- and instructor-generated, annotated sketches for learning, teaching, and assessment in geology courses. *Journal of Geoscience Education*, *53*, 85–95.

Kali, Y., & Orion, N. (1996). Spatial abilities of high-school students in the perception of geologic structures. *Journal of Research in Science Teaching*, *33*, 369–391.

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, *38*, 1008–1022.

Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computer-supported, case-based worked examples: Effects of erroneous examples and feedback. *Medical Education*, *42*(8), 823–829.

Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79–97.

Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning & Instruction*, *22*, 16–26.

Leutner, D., & Schmeck, A. (2014). The generative drawing principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 433–448). Cambridge, England: Cambridge University Press.

McLaren, B. M., Van Gog, T., Ganoe, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in computer-based learning environments. *Computers in Human Behavior*, *55*, 87–99.

McTigue, E., & Croix, A. (2010). Visual literacy in science. *Science Scope*, *33*, 17–22.

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*, 465–489.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. *Cognitive Models of Science*, *15*, 3–44.

Newcombe, N. S., & Stieff, M. (2012). Six myths about spatial thinking. *International Journal of Science Education*, *34*(6), 955–971.

Nievelstein, F., Van Gog, T., Van Dijck, G., & Boshuizen, H. P. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology*, *38*(2), 118–125.

Ohlsson, S. (1996). Learning from error and the design of task environments. *International Journal of Educational Research*, *25*, 419–448.

Ormand, C. J., Manduca, C., Shipley, T. F., Tikoff, B., Harwood, C. L., Atit, K., & Boone, A. P. (2014). Evaluating geoscience students' spatial thinking skills in a multi-institutional classroom study. *Journal of Geoscience Education*, *62*, 146–154.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.

Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In J. L. Plass, R. Moreno, & R. Brucken (Eds.), *Cognitive load theory* (pp. 89–108). Cambridge, England: Cambridge University Press.

Riggs, E. M., Lieder, C. C., & Balliet, R. (2009). Geologic problem solving in the field: Analysis of field navigation and mapping by advanced undergraduates. *Journal of Geoscience Education*, *57*(1), 48–63.

Roth, W. M., Bowen, G. M., & McGinn, M. K. (1999). Differences in graph-related practices between high school biology textbooks and scientific ecology journals. *Journal of Research in Science Teaching*, *36*, 977–1019.

Scheiter, K., Schleinschok, K., & Ainsworth, S. (2017). Why sketching may aid learning from science texts: Contrasting sketching with written explanations. *Topics in Cognitive Science*, *9*, 866–882.

Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology*, *39*, 275–286.

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology*, *102*, 872–879.

Shipley, T. F., Tikoff, B., Ormand, C., & Manduca, C. (2013). Structural geology practice and learning, from the perspective of cognitive science. *Journal of Structural Geology*, *54*, 72–84.

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). New York, NY: Cambridge University Press.

Slough, S. W., McTigue, E. M., Kim, S., & Jennings, S. K. (2010). Science textbooks' use of graphical representation: A descriptive analysis of four sixth grade science texts. *Reading Psychology*, *31*(3), 301–325.

Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap*. New York, NY: Simon & Schuster.

Suwa, M., & Tversky, B. (1997). What do architects and students perceive in their design sketches? A protocol analysis. *Design Studies*, *18*, 385–403.

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

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

Turner, S., & Libarkin, J. (2012). Novel applications of tablet PCs to investigate expert cognition in the geosciences. *Computers & Geosciences*, *42*, 162–167.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how? In *Psychology of learning and motivation* (pp. 147–181). San Diego, CA: Academic Press.

Van Gerven, P. W. M., Paas, F. G. W. C., Van Merriënboer, J. J. G., & Schmidt, H. G. (2002). Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning & Instruction*, *12*, 87–105.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*(3), 212–218.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*, 285–325.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535–585.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817–835.

Wernecke, U., Schütte, K., Schwanewedel, J., & Harms, U. (2018). Enhancing conceptual knowledge of energy in biology with incorrect representations. *CBE-Life Sciences Education*, 17. https://doi.org/10.1187/cbe.17-07-0133

White, R., & Gunstone, R. (1992). *Probing understanding*. New York, NY: Routledge.

Winn, W. D. (1994). Contributions of perceptual and cognitive processes to the comprehension of graphics. In W. Schnotz & R. Kulhavy (Eds.), *Comprehension of graphics* (pp. 3–27). Amsterdam, the Netherlands: Elsevier.

Wu, H. K., Krajcik, J. S., & Soloway, E. (2001). Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom. *Journal of Research in Science Teaching*, *38*, 821–842.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition & Instruction*, *4*, 137–166.

How to cite this article: Jaeger AJ, Marzano JA, Shipley TF. When seeing what's wrong makes you right: The effect of erroneous examples on 3D diagram learning. *Appl Cognit Psychol*. 2020;34:844–861. https://doi.org/10.1002/acp.3671

APPENDIX A.

Geologic block cross-sectioning test.

Appendix A

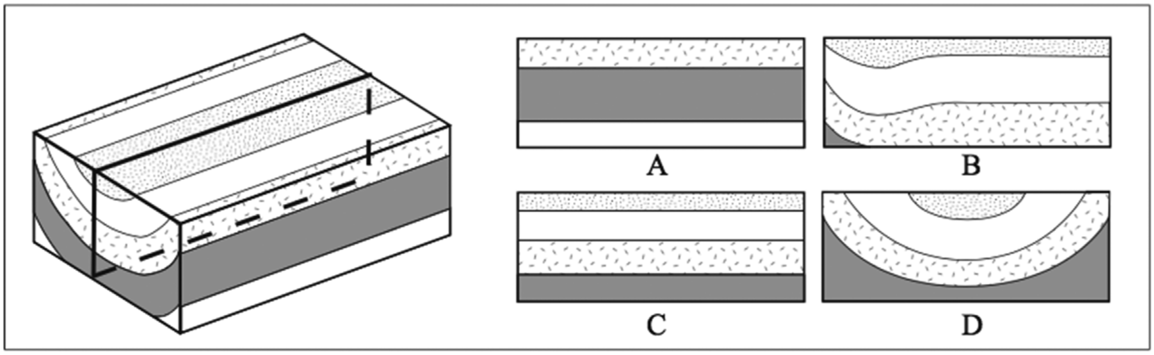
Geologic Block Cross-sectioning Test

This test is designed to assess your ability to mentally slice through a three dimensional geologic structure expressed in a block diagram.

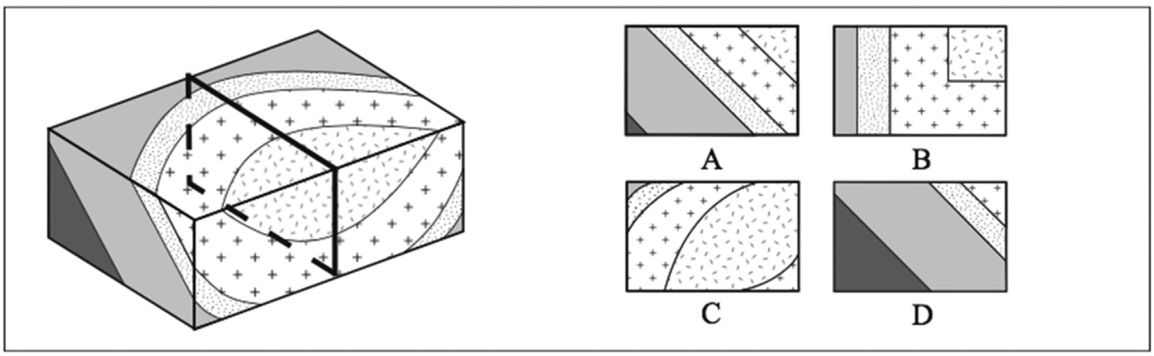
For each item below:

1. Study the geologic structure that is displayed in the 3-D block diagram.
2. Visualize what the cross-section of that geologic structure would look like on the surface of the vertical plane intersecting the block.
3. Choose the multiple choice answer that illustrates the structure along that plane. Where more than one answers appears to be possible, choose the MOST LIKELY answer.

Here is an example:



The answer to this example problem is **C**. It shows the layers in the correct positions, with the correct thicknesses, and in the correct orientations. Here is a second example:

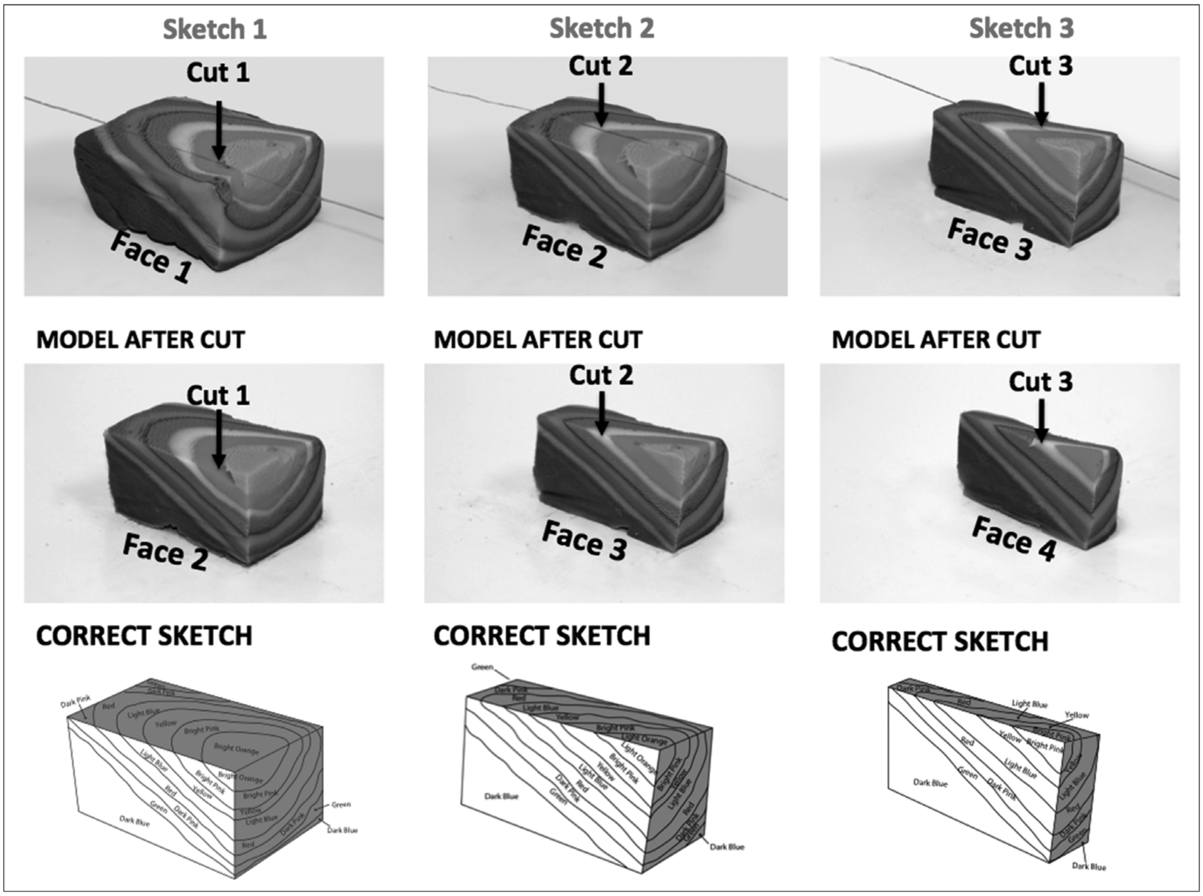


The answer to this example problem is **A**. It shows the layers in the correct positions, with the correct thicknesses, and in the correct orientations.

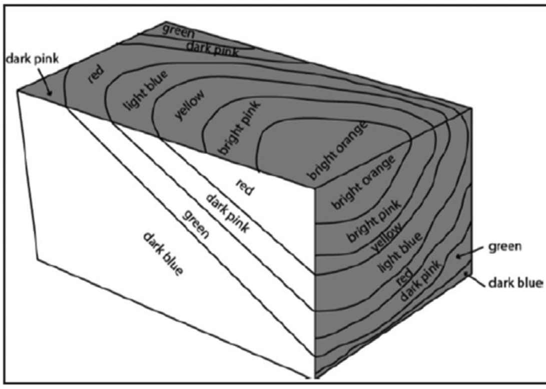
APPENDIX B.

Geologic block 1 showing three cuts and cross-sections, and three correct sketches.

Geologic Block 1 showing three cuts and cross-sections, and three correct sketches.



APPENDIX C.

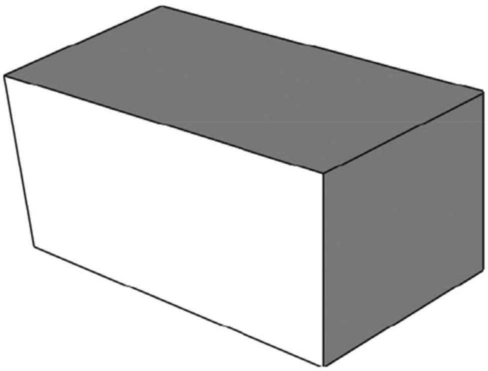


Part 2: Find Errors
Mark the differences between this sketch and the correct sketch

Part 3: Explanation
How was the student's sketch incorrect on the perpendicular side?

How was the student's sketch incorrect on the top side?

How was the student's sketch incorrect on the front face?



Part 2: Explanation
How did you use the layers visible on the perpendicular side to make your copy/sketch?

How did you use the layers visible on the top side to make your copy/sketch?

How did you use the layers visible on the front face to make your copy/sketch?

Part 3: Find Errors
Mark the differences between your copy/sketch and the correct sketch