Instructional Science
https://doi.org/10.1007/s11251-023-09630-y

ORIGINAL RESEARCH

Teaching advanced surgical anatomy with visual representations: comparing perceptual fluency and sense making

Christopher C. Stahl[1,2] · **Martina A. Rau**[1] · **Jacob A. Greenberg**[3]

Received: 15 July 2022 / Accepted: 15 March 2023

Abstract

Combinations of perceptual fluency and sense-making competencies contribute synergistically to learning gains in undergraduate science, technology, engineering, and math (STEM) education. However, instructional principles depend on the target of instruction, and in many fields, the targets of instruction are quite different from undergraduate STEM education. Professional learning often involves the application of previously acquired conceptual knowledge in a perceptually complex reality. This paper focuses on the field of surgery, specifically the recognition of surgical anatomy, in which the target of instruction is perceptual ability rather than conceptual knowledge. We conducted two experiments in which 42 and 44 surgical trainees participated in perceptual-fluency and sense-making interventions, followed by tests of their ability to recognize surgical anatomy in real operative images. The results showed that perceptual-fluency interventions contributed to gains in perceptual knowledge relating to surgical anatomy, whereas sense-making interventions did not. We discuss our findings in terms of alignment between instructional design and instructional goals, and the application of advances in learning sciences to adult learning of complex skills.

Keywords Adult learning · Perceptual fluency · Sense making · Visual representations · Medical education

Introduction

Visual representations can enhance learning (Ainsworth, 1999, 2006; Bodemer & Faust, 2006). However, students often find it difficult to work with visual representations (Ainsworth et al., 2002). Prior research in undergraduate science, technology, engineering, and

✉ Christopher C. Stahl
cstahl@uwhealth.org

[1] Department of Educational Psychology, University of Wisconsin-Madison, Madison, WI, USA

[2] Department of Surgery, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, Madison, WI 53792, USA

[3] Department of Surgery, Duke Health, Durham, NC, USA

Published online: 20 April 2023

math (STEM) education shows that providing instructional support for students' representational competencies enhances their learning of content knowledge (Rau et al., 2017b; Rau & Wu, 2018). Specifically, students need support for the ability to make sense of visual representations (sense making) as well as for the ability to fluently perceive meaning in visual representations (perceptual fluency) (Rau et al., 2017b; Rau & Wu, 2018). What is more, the ability to make sense of visual representations seems to be a prerequisite for students' benefit from perceptual-fluency interventions (Rau & Wu, 2018; Rau et al., 2017b).

However, in many domains, particularly outside of the school setting, the goal is not solely to support content learning. In the medical domain alone, expertise in fields such as radiology, surgery, dermatology, and pathology involves perceptual fluency of complex visual stimuli (e.g., diagnosing tumors in radiographic images). Therefore, unlike in undergraduate STEM education, visual representations do not merely serve as a tool to convey content knowledge. Rather, perceptual fluency itself should be a major target of medical instruction.

This difference between undergraduate STEM education and medical domains such as surgery is important because the target of instruction determines which instructional principles apply to the task at hand. Specifically, Koedinger and colleagues (2012) propose the *alignment hypothesis*: instructional principles do not apply universally but depend on the type of knowledge that is the target of instruction. A different target of instruction (i.e., perceptual fluency rather than content knowledge) may require different instructional principles than prior research in undergraduate STEM education has elucidated. Specifically, it is unclear whether sense making remains a prerequisite for students' benefit from perceptual-fluency interventions if the instructional target is perceptual in nature. Thus, testing the contribution of sense-making interventions and perceptual-fluency interventions to learning outcomes in domains with an inherent perceptual component will advance theory about learning with visual representations.

Addressing this question will also likely have practical implications for medical training. The focus of formal medical education is to impart content knowledge by supporting sense making with visual representations, whereas the role of perception is ignored (Kellman & Krasne, 2018). For example, surgery residents make sense of visual representations of human anatomy by studying verbal descriptions of images and photographs in textbooks and surgical atlases but do not receive instruction to improve perceptual fluency (Cadieux et al., 2021). This educational practice may be due to a tacit assumption that this perceptual expertise cannot be taught but can only be gained through experience. Thus, testing the effectiveness of sense-making and perceptual-fluency interventions may yield more effective approaches to medical training.

In this paper, we report on two experiments that used the perception of surgical anatomy as a target of instruction. Based on prior research in STEM, one would expect both sense-making and perceptual-fluency interventions to lead to increased gains, with the combination of the two interventions leading to higher learning gains than either intervention alone. Given the perceptual nature of surgery, Experiment 1 tested if the addition of sense-making to perceptual-fluency would lead to improved learning gains in perception. Participants were randomly assigned to receive (1) a sense-making intervention before a perceptual-fluency intervention or (2) only a perceptual-fluency intervention. The outcome of interest was the participants' ability to identify the inguinal hernia sac in surgical images.

Experiment 1 was limited in its ability to fully elucidate the impacts of the sense-making and perceptual-fluency interventions individually and only combined them in one sequence. Therefore, we conducted Experiment 2 to further investigate the relationship between sense making and perceptual fluency in teaching the perception of anatomy by using both possible sequences and performing more frequent testing to assess the impact of the interventions.

In Experiment 2, participants all underwent both a sense-making and a perceptual-fluency intervention, albeit in different orders. Learning gains were calculated after each intervention, allowing us to isolate the relative impact of each intervention. This allowed us to directly compare the efficacy of each intervention in isolation and check for both synergistic and sequencing effects of the interventional combinations. Thus, participants were randomized to receive (1) a sense-making intervention followed by a perceptual-fluency intervention or (2) a perceptual-fluency intervention followed by a sense-making intervention. The outcome of interest was participants' ability to identify the crura of the diaphragm in surgical images.

Our findings have implications for research on learning with multiple representations and for the design of instructional interventions in fields where the instructional target is perceptual in nature.

Literature review

Prior research suggests that instructional principles do not apply universally, but instead seem to depend on the complexity of the knowledge that is the target of instruction (Koedinger et al., 2012). Building on this observation, Koedinger and colleagues (2012) offer the *alignment hypothesis*, which suggests that the type of instruction needs to match the complexity of the learning goal. Additionally, according to Lowe et al. (1993), the appropriate selection of graphics to achieve desired learning outcomes requires empirical evidence, not just intuitive, superficial comparison. Domain-specific background knowledge also shapes the way individuals respond to visual-based tasks, so research should be performed with participants that would be targeted by the learning interventions (Lowe, 1993). Thus, research on instructional principles should be performed across a variety of fields with learning goals that vary in complexity. The present article focuses on a particular instructional principle; namely, that surgical students' learning with visual representations is enhanced if students receive support for both sense making and perceptual fluency. In the following, we review prior research that led to this principle. We then discuss how extant research might be extended by studying sense-making and perceptual-fluency interventions in the field of surgery, where the learning goal differs in complexity from undergraduate STEM education.

Sense making

Sense-making competencies allow students to explain what a visual representation shows (Rau, 2017). They involve the ability to explicitly, verbally describe the different components of a visual representation and relate them to abstract knowledge and concepts about the topic at hand (Ainsworth, 2006; Kozma & Russell, 2005). For example, a surgery resident looking at an illustrated anatomy atlas should recognize the red structures as arteries, the blue ones as veins, and the yellow ones as nerves, and they should connect these anatomic structures with conceptual knowledge about the structures themselves and which structures should be severed, manipulated, or avoided in an upcoming operation. Making such connections involves structure mapping because students need to map visual features of the representations to abstract concepts (Gentner, 1983; Schnotz, 2014).

Instructional interventions that support sense-making competencies (sense-making interventions for short) may prompt students to explain how visuals show a variety of structures and concepts, and often ask students to compare a variety of visual representations while reflecting on whether they show similar or different concepts (Ainsworth, 2006; Rau, 2017). These activities are typically designed to help learners focus on relevant visual features (Bodemer & Faust, 2006; Seufert, 2003).

Perceptual fluency

Much like one becomes fluent in a language—being able to use it without conscious thought—one can become fluent in perceiving meaning in a visual representation (Kellman et al., 2010). Experts are highly efficient at extracting meaningful information from visual representations with little cognitive effort, and they can integrate information distributed across multiple representations (Dreyfus & Dreyfus, 1986; Richman et al., 1996). Hence, perceptual fluency frees up limited mental resources for deeper reasoning and decision-making (Gobet & Clarkson, 2004; Goldstone et al., 1997, p. 199).

The processes through which students become perceptually fluent are qualitatively different from those through which students learn to make sense of visual representations. Perceptual fluency is acquired via implicit, nonverbal, inductive processes (Gibson, 2000; Goldstone et al., 1997; Kellman et al., 2010; Koedinger et al., 2012). These processes are often unintentional and seem to occur unconsciously (Frensch & Rünger, 2003). Verbal reasoning about the visual representations is not necessary and can even impede students' perceptual learning (Chin & Schooler, 2008; Kellman et al., 2010; Schooler et al., 1997). Rather, students induce patterns through exposure to many examples (Kellman & Massey, 2013; Rau, 2017).

Perceptual fluency can be deliberately supported by perceptual-fluency interventions (Amiri et al., 2020; Kellman et al., 2010; Krasne et al., 2013, 2020; Romito et al., 2016). These interventions offer numerous, short classification trials in which students receive simple feedback (i.e., correct/incorrect) (Kellman & Garrigan, 2009). These trials typically present many example visuals and provide extensive variation across trials, allowing students to train their perceptual system to identify variant and invariant properties of the examples (Kellman & Garrigan, 2009).

Most research on learning with visual representations has focused on sense-making competencies rather than perceptual fluency (Rau, 2017). This lack of focus on perceptual fluency may be due to a 'blind spot' that experts have for their own perceptual abilities (Rau et al., 2019). Automatic processes, such as the perception of meaning from visual representations, are difficult to access (Nathan et al., 2001). Yet, even if they are not aware of it, experts consistently perceive information more efficiently and effectively than novices. This effect has been established in numerous domains, highlighting the importance of perceptual fluency as a goal for learning (Chase & Simon, 1973; Chi et al., 1981; Egan & Schwartz, 1979; Gobet & Clarkson, 2004; Reitman, 1976; Strange & Jenkins, 1978; Zatorre, 1979). Consequently, recent research has investigated combinations of sense-making and perceptual-fluency interventions, reviewed in the following.

Combining sense-making and perceptual-fluency interventions

Sense-making and perceptual-fluency interventions are beneficial in isolation, but few studies have examined the interaction between these two types of interventions. Kellman and colleagues added perceptual-fluency interventions to math and chemistry instruction that was mostly focused on sense-making competencies and found benefits for students' learning of content knowledge (Kellman et al., 2008). Rau and colleagues provided experimental evidence that a combination of sense-making and perceptual-fluency interventions led to higher learning outcomes than either alone for undergraduate students learning about chemistry and math (Rau et al., 2017b; Rau & Wu, 2018). Further research revealed that students' prior knowledge moderates the order in which they should receive sense-making and perceptual-fluency interventions (Rau, 2018; Rau & Wu, 2018; Rau et al., 2017a).

While important, these findings are limited to learning content knowledge in math and chemistry fields. Content knowledge is arguably complex because it involves interpreting the visual representations in the context of conceptual information about domain-relevant topics (e.g., the equivalency of expanded fractions or atomic structure in chemistry). Yet, visual representations are critical adjuncts to learning in many other fields of study and populations where the learning goals focus on simpler knowledge types.

A further limitation regards the role of prior knowledge in moderating order effects of sense-making and perceptual-fluency interventions. The prior studies just reviewed focused on students who were learning material for the first time. However, in many professional training contexts, students have extensive educational experience with the targeted content but are learning to apply their knowledge in practical situations. Prior research has not established which order of sense-making and perceptual-fluency interventions is most effective in these contexts.

Our work seeks to expand our understanding of the importance of and interactions between sense-making and perceptual-fluency interventions in the field of surgery, a type of professional training for students who have received extensive medical education. In the following, we describe the learning goals around visual representations in this field and how their complexity differs from the learning goals in prior research.

Sense-making competencies and perceptual fluency in surgical anatomy recognition

The field of surgical anatomy is a suitable context for expanding prior research on representational competencies for several reasons. First, identifying surgical anatomy is fundamentally different from content knowledge in undergraduate STEM education in domains like math or chemistry. In undergraduate STEM education, learning is thought to involve the ability to make sense of domain-relevant concepts (Koedinger et al., 2012). Thus, tests of content knowledge that prior studies have used to assess conceptual knowledge of math and chemistry are much closer to sense-making interventions (designed to enhance students' ability to make sense of visuals) than they are to perceptual-fluency interventions (designed to lead to fluent perception of information from visuals). In that sense, the surprising finding of much of Rau's and Kellman's research is the contribution of perceptual-fluency interventions to students' performance on problems that primarily involved the ability to make sense of visual representations to reason conceptually about content knowledge.

In contrast, in many professional domains, including surgery, a key aspect of performance is the ability to perceive accurately—coined 'sensory semiosis' (Cope et al., 2015a, b). In domains like surgery, perceptual fluency is a main learning goal, not a means to an end that frees up cognitive resources for deeper conceptual reasoning. This difference from undergraduate STEM education leads to the reverse of the question addressed by the previous literature—instead of asking and ascertaining the boundary conditions by which perceptual-fluency interventions can enhance content learning, the goal of our research is to investigate whether adding a sense-making intervention to a perceptual-fluency intervention contributes to students' performance on problems that primarily involve perceptual knowledge.

This difference is noteworthy also because Koedinger and colleagues (2012) characterize perceptual knowledge as *simple* types of knowledge, whereas conceptual content knowledge and sense-making competencies are characterized as *complex* types of knowledge. Their *alignment hypothesis* predicts that simple types of interventions—such as perceptual-fluency interventions—are effective at enhancing students' acquisition of simple types of knowledge such as perceptual fluency. Further, Koedinger and colleagues (2012) suggest that simple interventions—such as perceptual-fluency interventions—can enhance learning of complex knowledge because even complex principles must be remembered and practiced, which is in line with the findings by Rau and Kellman described above. By contrast, Koedinger and colleagues (2012) argue that complex interventions are ineffective for simple types of knowledge. For instance, complex interventions such as sense-making interventions are ineffective for simple knowledge types such as grammar rules because there is nothing to make sense of when memorizing simple rules (Wylie et al., 2009). However, as mentioned, this has not been tested in contexts where the learning goal is perceptual in nature.

In sum, surgical anatomy is an ideal domain to study the interaction between sense-making and perceptual-fluency interventions. This work will extend our understanding of Koedinger and colleagues' (2012) alignment hypothesis by yielding new insights into the contributions of sense-making and perceptual-fluency competencies for learning perceptual knowledge. In addition to these theoretical implications, the findings of this research will have practical implications for surgical training. Despite the large amount of prior medical training participants in this study underwent (2–6 years), there is no dedicated training in anatomic perception. Currently, surgical anatomy instruction involves sense-making interventions with idealized representations, such as the study of textbooks and illustrated surgical atlases, with no formal instruction in perception (Cadieux et al., 2021). Even these sense-making interventions rarely focus on teaching how to identify actual anatomy, and simply label anatomic representations. Our findings will demonstrate whether perceptual-fluency interventions are effective for surgical training and (if so) how they should be combined with existing or novel sense-making interventions.

Experiment 1

To address the limitations of prior research just reviewed, Experiment 1 addresses the following research question (RQ):

RQ1 Does the addition of a sense-making intervention to a perceptual-fluency intervention improve perception of surgical anatomy?

Methods

Participants

Participants were 42 medical students and surgery residents from a single academic medical center with a moderate-sized general surgery residency (7 residents/year). They

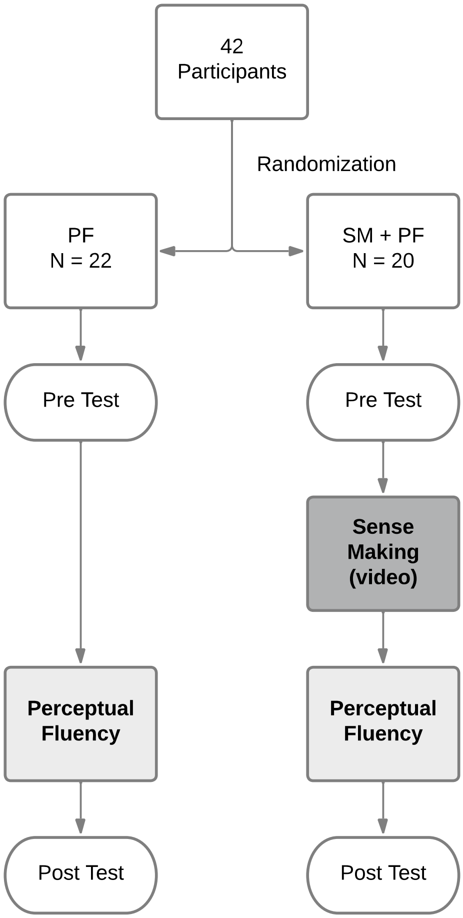

Fig. 1 Experiment 1 design. Shaded boxes represent interventions. *PF* perceptual-fluency intervention, *SM* sense-making intervention

were recruited by email solicitation. Sixteen participants were medical students (0–4 months of clinical surgical experience), while the remaining 26 were residents (1–5 years of clinical surgical experience). The experiment was approved by the Institutional Review Board (IRB 2019–1328).

Experimental design

The experiment compared two conditions, illustrated in Fig. 1. The *perceptual-fluency only condition* (PF-only condition) received only the perceptual-fluency intervention. The *sense-making and perceptual-fluency condition* (SM + PF condition) received the sense-making intervention followed by the perceptual-fluency intervention. Participants were block-randomized to experimental conditions stratified by experience level (high experience: residents post-graduate year 3 to 5; low experience: medical students up to post-graduate year 2 residents). Of the 42 participants, 22 participants were in the PF-only condition, while 20 participants were in the SM + PF condition.

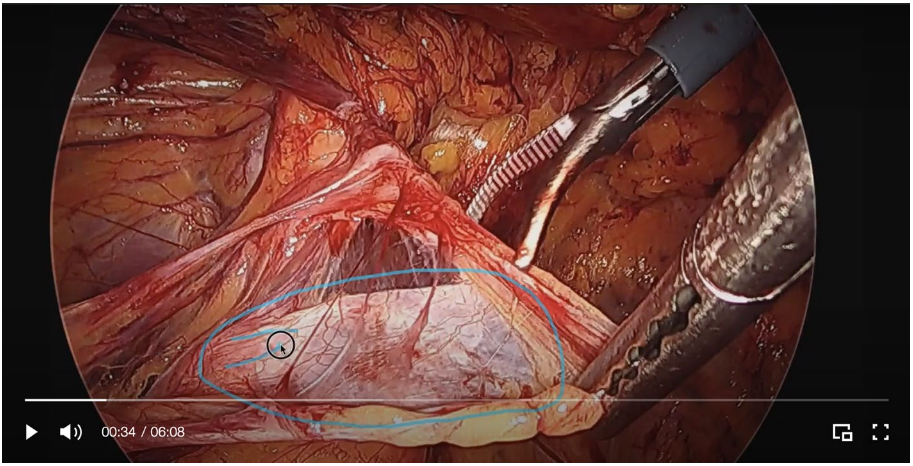

Fig. 2 Sense-making intervention: participants watched a 6-minute video

Materials

The interventions were created and administered as interactive websites (Wix.com Ltd., Tel Aviv, Israel). Participants accessed the websites via email links at a computer, time, and place of their choosing. The websites were password-protected to prevent unauthorized use, and each participant was given a unique confidential identifier to associate with their results. The interventions provided instruction on the recognition of the hernia sac (an outpouching of the inner peritoneal lining of the abdomen into the inguinal canal) during a laparoscopic inguinal hernia repair.

During the sense-making intervention, participants watched a 6-minute video discussing visual cues that can identify the hernia sac from the surrounding structures in the operative field. The video contained still images from laparoscopic inguinal hernia repairs (similar to but distinct from those used in testing), with corresponding narration and on-screen drawing highlighting the hernia sac and the visual cues present in the operative field. Figure 2 shows a screenshot of the video.

In the perceptual-fluency intervention, participants were given unlimited time to work through 30 “visual flashcards” showing images of laparoscopic inguinal hernia repairs containing the hernia sac within the operative field. These flashcards were created using still images captured from videos of laparoscopic inguinal hernia repairs. The images were converted into polygonal html image maps, with the polygon overlying the hernia sac. Participants were shown an image and instructed to click on the hernia sac. They received immediate feedback on whether their selection was correct or incorrect and were shown an outline of the correct response (see Fig. 3).

Measures

Tests We assessed participants’ acquisition of perceptual knowledge with equivalent test versions; that is, they included items with images of similar difficulty that contained similar perceptual distractors and comparable views of the operative anatomy. The two test versions

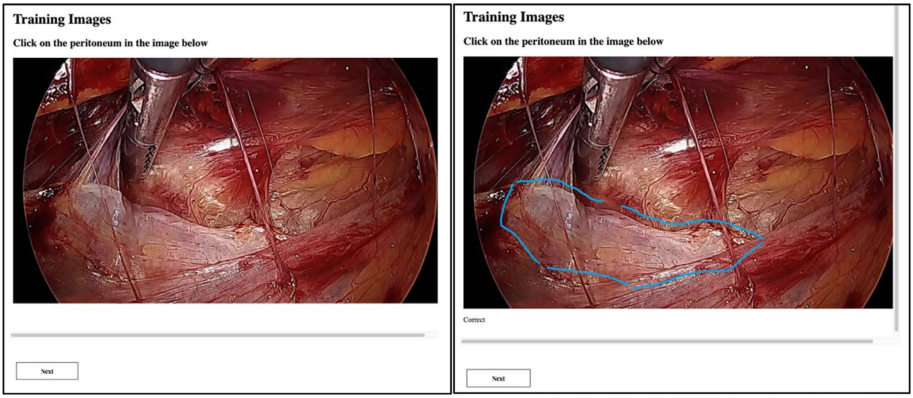

Fig. 3 Perceptual-fluency intervention: after clicking, participants were shown an outline of the correct structure and given feedback on whether their selection was correct or incorrect

were administered as pre-test and post-test (see Fig. 1) and counterbalanced across these test times to prevent possible effects of the different test versions. Each test consisted of ten items. Each item consisted of an image of laparoscopic inguinal hernia repairs containing the hernia sac. Participants were instructed to click on the hernia sac. Once clicked, the item was marked as complete, but no feedback was given. All tests contained unique images that were not shown in any other test or any of the perceptual-fluency interventions. The pre-test consisted of ten items and the post-test of 20 items.

We confirmed the validity of the tests in two ways. First, we administered the pre-test to three attending surgeons with > 15 years of experience in minimally invasive surgery. These experts scored considerably better on the pre-test than the participants, with an average score of 0.90, confirming that this test measures a skill that improves with surgical expertise. Second, pre-test scores for participants were stratified by experience level. Participant pre-test scores increased with increasing amounts of clinical surgical training (< 1 year: 0.42, 1–2 years: 0.56, 3–5 years: 0.84, > 15 years: 0.90). The test was reliable (Cronbach's alpha = 0.86).

Survey After completion of the study, we administered an online survey (Qualtrics, Provo, UT, USA) to the participants. The survey consisted of thirteen questions. The survey confirmed participants' clinical experience, asked about subjective experience with the interventions, participants' perception of learning, and gathered information on baseline study habits. Further demographic data was not collected to preserve the anonymity of the participants given the small class sizes for a given post-graduate year.

Procedure

Participants participated in the experiment online. First, all participants watched a 2-minute video explaining the testing procedures and website navigation. As shown in Fig. 1, they then took a pre-test, followed by the interventions they were assigned to (PF vs. SM + PF). Then, they took the post-test. The survey was administered after the post-test.

Table 1 Means (standard deviations) for dependent measures by condition

Measure	PF-only	SM+PF
Post-graduate years	2.182 (1.868)	1.900 (1.889)
Pre-test	0.564 (0.261)	0.580 (0.322)
Post-test	0.811 (0.108)	0.803 (0.091)

Analyses

To investigate whether the sense-making intervention contributes to learning of perceptual knowledge (RQ1.1), we used a repeated measures ANOVA with condition as an independent between-subjects factor, test-time (pre-test, post-test), as a repeated, within-subjects factor, and scores on the tests as dependent measures. We examined the interaction between test time and condition. We report partial η^2 for effect sizes. An effect size of 0.01 corresponds to a small effect, 0.06 to a medium effect, and 0.14 to a large effect (Cohen, 1988).

Results

All participants completed the experiment, yielding a final sample of $N=42$. However, only 37 participants completed the survey. Table 1 shows the descriptives by condition.

Power analysis

To assess the sensitivity of our analyses, we conducted a post hoc power analysis. Our ANCOVA model with an assumed α error probability of 0.05, power of $(1\text{-}\beta)=0.80$, and $N=42$ was able to detect large effect sizes of p. $\eta^2=0.164$ or larger. In light of prior research that showed medium to large effects on perceptual fluency in the medical field ranging between $d=0.8$ to $d=1.6$ (Krasne et al., 2013), our study seems to be sufficiently powered.

Prior checks

First, Levene's test revealed no evidence that the assumption of normal distribution was violated for the dependent measures (ps > 0.10).

Second, we checked for differences between conditions prior to the intervention using a MANOVA with condition as the independent factor and post-graduate years and pre-test scores as dependent measures. We found no differences between conditions in terms of post-graduate years ($F<1$, p. $\eta^2=0.006$) and pre-test scores ($F<1$, p. $\eta^2=0.001$).

Third, we checked that participants exhibited learning gains. To this end, we examined the effect of test time in the ANOVA model described above (see "Analyses" section). Results showed a large significant effect of test-time, $F(1, 40)=33.881$, $p<.001$, p. $\eta^2=0.459$, such that participants' scores improved from pre-test to post-test.

Condition effects

To address RQ1.1 (whether the sense-making intervention contributes to learning perceptual knowledge), we examined the interaction between test time and condition in the ANOVA model described above (see "Analyses" section). We found no significant effect of condition ($F < 1$, p. $\eta^2 = 0.002$), suggesting that the more sense-making intervention did not enhance learning of perceptual knowledge.

Participant experience with the educational interventions

Most participants (60%, $n = 22$) felt that the test questions were "somewhat difficult" or "extremely difficult". Subjectively, most participants (84%, $n = 31$) agreed or strongly agreed with the statement "I improved my ability to see surgical structures as a result of the study activities". All participants stated they would either "definitely" or "probably" use interventions like our perceptual-fluency intervention to study surgical anatomy if available. A MANOVA showed no differences between conditions on these questions ($ps > 0.30$).

Discussion

Experiment 1 revealed no differences between conditions on the post-test. This suggests that the sense-making intervention did not contribute to learning to perceive surgical anatomy. However, Experiment 1 has several limitations. For instance, participants in the SM + PF condition received the sense-making intervention before the perceptual-fluency intervention. While this choice was based on prior studies showing that this sequence is most effective (Rau et al., 2017; Rau, 2018), these prior studies were conducted in the context of undergraduate STEM education. As argued above, there are qualitative differences between STEM domains and the medical domain of surgical anatomy that our experiment focused on—and importantly, in this prior work sequencing effects have been demonstrated. This raises the question of whether the sense-making intervention might be beneficial if it is provided after the perceptual-fluency intervention in the context of surgical anatomy recognition. Experiment 2 formally addresses this question.

A further limitation of Experiment 1 is that the sense-making intervention was presented in the form of videos. As argued above, surgical anatomy instruction often involves sense-making of visual representations in textbooks and illustrated surgical atlases. Based on this observation, the sense-making intervention in Experiment 2 uses a more conventional design.

Experiment 1 also did not assess the durability of learning gains. An effective learning intervention should have durable results over time. Sense-making processes may facilitate more durable learning gains due to explicitly created mental rules and heuristics that can be recalled, unlike unconscious mental processes. Experiment 2 incorporates a delayed test of perceptual knowledge to assess learning gains over time.

A final limitation of Experiment 1 is that it did not measure nor hold time on task constant across conditions. Participants in the SM + PF condition received additional instructional time due to the inclusion of the sense-making intervention, compared to participants in the PF-only condition. Our finding that the sense-making intervention did not enhance learning outcomes is particularly noteworthy because this shows that the added instructional time in the SM + PF condition did not pay off. Nevertheless, Experiment 2 employs an experimental design that records time on task across conditions, which allows for this variable to be accounted for in the statistical analysis.

Experiment 2

To address the limitations of Experiment 1, Experiment 2 addresses the following research questions (RQs):

RQ2.1 What are the relative contributions of sense-making vs. perceptual-fluency interventions when learning perceptual knowledge?

RQ2.2 Does the order of sense-making and perceptual-fluency interventions affect learning of perceptual knowledge?

Methods

Unless otherwise noted, the methods for Experiment 2 were identical to Experiment 1.

Participants

Participants were 44 medical students and surgery residents recruited from the same academic medical center. Twelve participants were medical students (0–4 months of clinical surgical experience), while the remaining 32 were residents (1–5 years of clinical surgical experience). Due to low numbers of eligible participants, subjects that had completed Experiment 1 were eligible to participate in Experiment 2. We did not anticipate any carry over effects from Experiment 1 due to the unique content in the second experiment. The experiment was approved by the Institutional Review Board (IRB 2019–1328).

Experimental design

The experiment used a cross-over design with two conditions (see Fig. 4). Participants either received a perceptual-fluency intervention followed by a sense-making intervention (PF-first condition), or they received a sense-making intervention followed by a perceptual-fluency intervention (SM-first condition). Participants were block-randomized to experimental conditions using the same procedure as in Experiment 1. Of the 44 participants, 23 were in the PF-first condition, while 21 were in the SM-first condition.

Materials

The interventions were created and administered as in Experiment 1, except that the interventions provided instruction on recognition of the crura of the diaphragm (i.e., muscular bands of the diaphragm that surround the esophagus at the esophageal hiatus where it enters the peritoneal cavity) during a laparoscopic hiatal hernia repair.

In the sense-making intervention, participants were given unlimited time to review five representative pages of a surgical textbook. As shown in Fig. 5, the pages

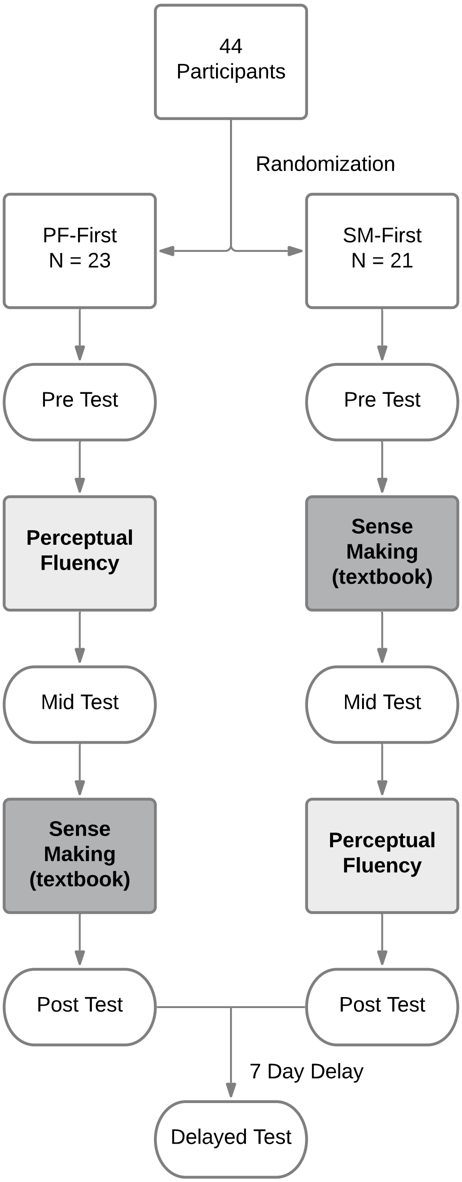

Fig. 4 Experiment 2 design. Shaded boxes represent interventions. *PF* perceptual-fluency intervention, *SM* sense-making intervention

contained text descriptions of a hiatal hernia repair along with multiple labeled anatomic illustrations and operative images. These visual representations depicted the anatomy of the diaphragm, esophagus, and stomach.

The perceptual-fluency intervention was identical to Experiment 1, except that they included images of laparoscopic hiatal hernia repairs (see Fig. 6).

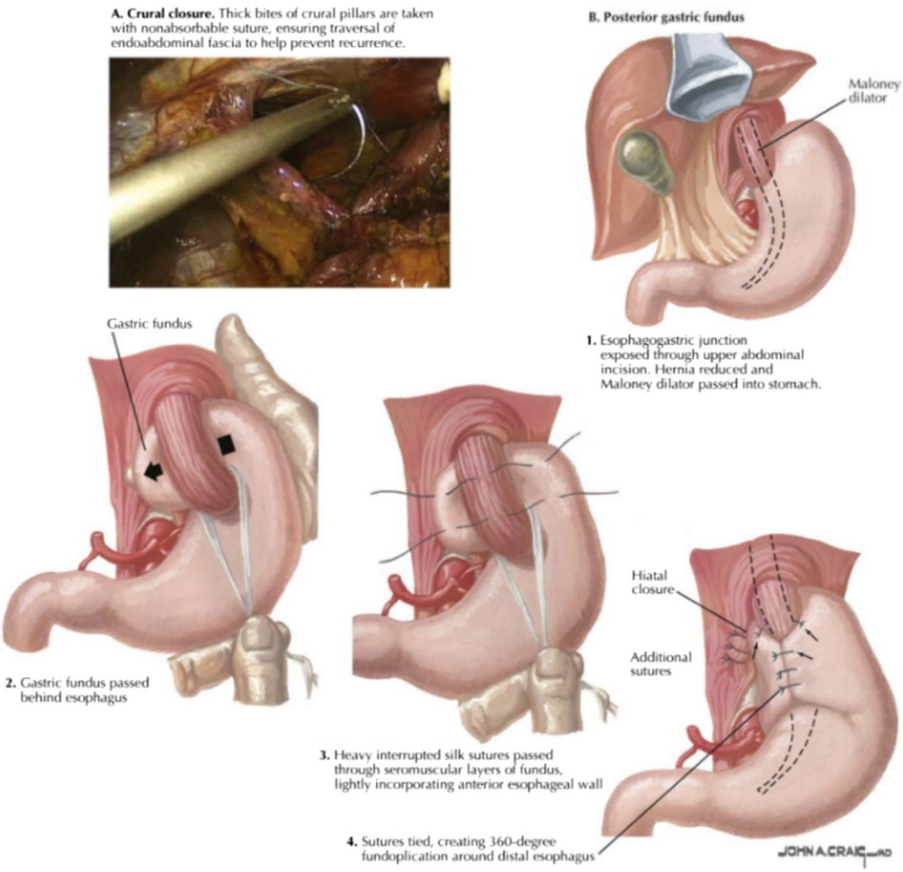

Fig. 5 Sense-making intervention: participants viewed pages from a surgical textbook

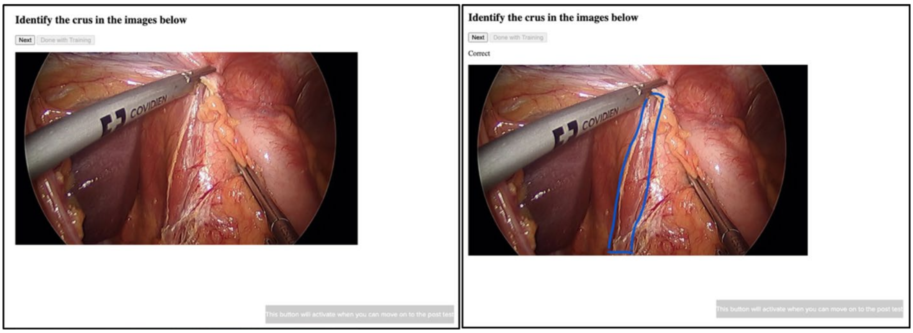

Fig. 6 Perceptual-fluency intervention: after clicking, participants were shown an outline of the correct structure and given feedback on whether their selection was correct or incorrect

Measures

Tests We assessed participants' acquisition of perceptual knowledge with equivalent test versions administered as a pre-test, mid-test, post-test, (all administered on the same day), and delayed-test (administered 1 week later) (see Fig. 4). These tests were identical in structure to the tests described in Experiment 1, but they used images of laparoscopic hiatal hernia repairs containing the crus, rather than laparoscopic inguinal hernia repairs containing the hernia sac. Each test consisted of ten items, each of which showed an image of a laparoscopic hiatal hernia repair.

As in Experiment 1, we confirmed the validity of the tests in two ways. First, we administered the test to three attending surgeons with over 15 years of experience in minimally invasive surgery. These experts scored considerably better than any of the participants on the pre-test, with an average score of 0.90, confirming that this test measures a skill that improves with surgical expertise. Second, pre-test scores for participants were stratified by experience level. Participant pre-test scores increased with increasing amounts of clinical surgical training (<1 year: 0.32, 1–2 years: 0.50, 3–5 years: 0.59, > 15 years: 0.90). The tests were reliable (Cronbach's alpha = 0.78).

Survey After completion of the study, participants took a survey like the one in Experiment 1.

Time on task We calculated time on task as the time spent on the perceptual-fluency intervention (PF-time) and the sense-making intervention (SM-time) by subtracting timestamps corresponding to start and finish transactions on each intervention.

Procedure

Participants participated in the experiment online. They first watched a 2-minute video explaining the testing procedures and website navigations. As shown in Fig. 4, they then took a pre-test, followed by an intervention (perceptual-fluency or sense-making) and a mid-test. They then crossed over to the other intervention (sense-making or perceptual-fluency), and after completion took a post-test. Finally, all participants took a delayed test 7 days after the post-test to assess the durability of any learning gains. De-identified data including test scores, training scores, and time on task were automatically collected via online databases. The survey was administered following the post-test.

Analyses

To gain insights into the unique benefits of the sense-making intervention vs. the perceptual-fluency intervention (RQ2.1), we used planned contrasts to compare participants' scores on the mid-test. To investigate whether the order of sense-making and perceptual-fluency interventions affects learning of perceptual knowledge (RQ2.2), we used a repeated measures ANCOVA with test-time (i.e., mid-test, post-test, and delayed-test) as repeated within-subjects factor, condition as an independent between-subjects factor, participants' pre-test scores as a covariate, and test scores as dependent measures. In all analyses, family-wise type I error rates were controlled at $\alpha = 0.05$ using Bonferroni adjustment.

Table 2 Means (standard deviations) for dependent measures by condition

Measure	SM-first	PF-first
Post-graduate years	2.048 (1.717)	2.348 (1.824)
Pre-test	0.438 (0.277)	0.513 (0.230)
Mid-test	0.521 (0.202)	0.789 (0.120)
Post-test	0.879 (0.123)	0.837 (0.083)
Delayed-test	0.795 (0.127)	0.789 (0.105)
Time on sense-making intervention (min)*	5.98 (4.95, 8.05)	3.18 (1.72, 4.84)
Time on perceptual-fluency intervention (min)	4.607 (3.221)	3.420 (0.951)

*Median and interquartile range reported due to large standard deviation

Results

All 44 participants completed the pre-, mid-, and post-tests, and survey, but six participants were lost to follow-up (did not respond to 3x email follow-up communications for the delayed test) before completing the delayed test (four in the PF-first condition, two in the SM-first condition). These participants were excluded from the analyses, yielding a final sample of $N = 38$. Table 2 shows the descriptives by condition.

Power analysis

To assess the sensitivity of our analyses, we conducted a posthoc power analysis. Our repeated measures ANCOVA model with an assumed α error probability of 0.05, power of $(1\text{-}\beta) = 0.80$, $N = 38$, and an average correlation among repeated measures of $r = .211$ was able to detect medium effect sizes of p. $\eta^2 = 0.065$ or larger.

Prior checks

First, Levene's test revealed no evidence that the assumption of normal distribution was violated for the dependent measures (ps > 0.10).

Second, we checked for differences between conditions prior to the intervention using a MANOVA with condition as the independent factor and post-graduate years, pre-test scores, SM-time, and PF-time as dependent measures. We found no differences between conditions in terms of post-graduate years ($F < 1$) and pre-test scores ($F < 1$). Participants across conditions spent similar amounts of time on the perceptual-fluency intervention, $F(1, 42) = 2.857$, $p = .098$. However, participants in the SM-first condition spent more time on the sense-making intervention than participants in the PF-first condition, $F(1, 42) = 4.776$, $p = .034$, p. $\eta^2 = 0.102$. Because time spent on the sense-making intervention correlates significantly with scores on the mid-test ($r = -.312$, $p = .039$), we conducted the following analyses with and without SM-time as a covariate.

Third, we checked that participants exhibited learning gains. To this end, we performed a repeated measures ANCOVA with test-time (pre-, mid-, post-, delayed-test) as repeated

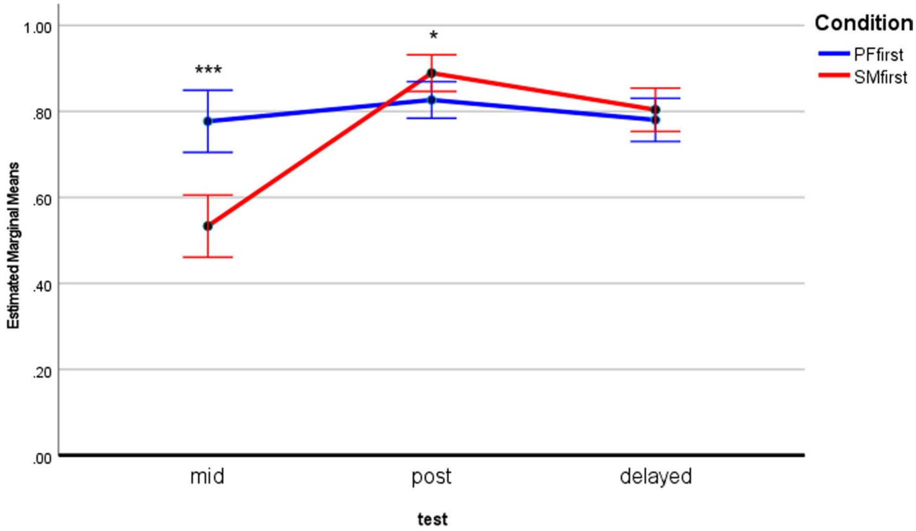

Fig. 7 Experiment 1 results by condition after controlling for pre-test scores. Conditions received their first intervention between the pre-test and the mid-test, and the second intervention between mid-test and post-test. Error bars represent 95% confidence intervals. ***$p<.001$, **$p<.01$, *$p<.05$; adjusted by Bonferroni correction

within-subjects factor, condition as the independent factor, and test scores as the dependent measures. Results showed a large significant effect of test-time, $F(3, 108)=57.669$, $p<.001$, p. $\eta^2=0.616$. Predefined contrasts showed that participants' scores on the mid-test, post-test, and delayed-test were significantly higher compared to the pre-test (ps <0.001).

Finally, we checked whether sphericity can be assumed for the repeated ANCOVA model we used for RQ 2.1 using Greenhouse-Geisser, Huynh-Feldt, and lower-bound ε estimates. We found no indication that sphericity was violated.

Condition effects

To address RQ2.2 (whether the order of sense-making and perceptual-fluency interventions affects learning of perceptual knowledge), we used the repeated ANCOVA model described above. Results showed a medium-sized significant effect of condition, $F(1, 35)=4.857$, $p=.034$, p. $\eta^2=0.122$, which was qualified by a large significant interaction between condition and test-time, $F(2, 70)=18.302$, $p<.001$, p. $\eta^2=0.344$, illustrated in Fig. 7[1].

To gain insights into the interaction, we performed post hoc pairwise comparisons of the conditions at each test time. Relevant to RQ2.2 are the effects at the post-test and the delayed-test; that is after both conditions received both interventions. Results showed a

[1] When including SM-time as a covariate, the main effect of condition was no longer significant, $F(1, 34)=3.238$, $p=.081$, p. $\eta^2=0.087$; but the interaction between condition and test-time remained significant, $F(2, 70)=14.870$, $p<.001$, p. $\eta^2=0.304$.

medium-sized significant advantage of the SM-first condition at the post-test, $F(1, 35) = 4.377$, $p = .044$, p. $\eta^2 = 0.111$[2]. However, there was no significant effect of condition at the delayed-test ($F < 1$)[3]. These results suggest that providing the sense-making intervention first had a slight advantage, which was, however, short-lived because it washed out after a 7-day delay.

We further explored the unique benefit of the sense-making intervention and the perceptual-fluency intervention using post hoc pairwise comparisons at the mid-test (RQ2.1). Recall that the mid-test was given after participants in the SM-first condition had received the sense-making intervention but not the perceptual-fluency intervention; whereas the PF-first condition had received the perceptual-fluency intervention but not the sense-making intervention. Results showed a large significant advantage of the PF-first condition at the mid-test, $F(1, 35) = 22.990$, $p < .001$, p. $\eta^2 = 0.396$[4].

To better understand this effect, we used within-condition comparisons. As shown in Fig. 8, each condition exhibited significant learning gains only directly after receiving the perceptual-fluency intervention; that is, the PF-first condition showed learning gains at the mid-test, $F(1, 18) = 37.275$, $p < .001$, p. $\eta^2 = 0.674$[5], but no learning gains at any other test-time (ps > 0.05)[6].

The SM-first condition showed no significant learning gains at the mid-test; that is, before having received the perceptual-fluency intervention, $F(1, 18) = 2.177$, $p = .157$[7]. However, the SM-first condition showed large learning gains after receiving the perceptual-fluency intervention; that is, at the post-test, $F(1, 18) = 60.313$, $p < .001$, p. $\eta^2 = 0.770$[8], as well as at the delayed-test, $F(1, 18) = 9.143$, $p = .007$, p. $\eta^2 = 0.337$[9]. In sum, these patterns of results suggest that the perceptual-fluency intervention, but not the sense-making intervention, contributed to gains in perceptual knowledge.

Participant experience with the interventions

The majority of participants felt that the tests were "somewhat difficult" or "extremely difficult" (70%, $n = 31$), highlighting the lack of preparation for such perceptual tasks in medical education. An ANOVA showed that participants in the SM-first condition reported that they found the tests more difficult, compared to participants in the PF-first condition, $F(1, 42) = 5.464$, $p = .024$, p. $\eta^2 = 0.115$. When asked to compare the perceptual-fluency and sense-making interventions, most participants preferred the perceptual-fluency intervention (73%, $n = 32$). A chi-square test showed no differences between conditions on this question. Most participants stated that they would "definitely yes" or "probably yes" incorporate a

[2] When including SM-time as a covariate, the effect remained significant, $F(1, 34) = 4.817$, $p = .035$, p. $\eta^2 = 0.123$.

[3] When including SM-time as a covariate, the result was the same ($F < 1$).

[4] When including SM-time as a covariate, the effect remained significant, $F(1, 34) = 18.081$, $p < .001$, p. $\eta^2 = 0.347$.

[5] When including SM-time as a covariate, the effect remained significant, $F(1, 17) = 10.311$, $p = .005$, p. $\eta^2 = 0.378$.

[6] When including SM-time as a covariate, the results were the same (ps > 0.05).

[7] When including SM-time as a covariate, the results were the same, $F(1, 17) = 3.414$, $p = .082$.

[8] When including SM-time as a covariate, the effect remained significant, $F(1, 17) = 29.273$, $p < .001$, p. $\eta^2 = 0.633$.

[9] When including SM-time as a covariate, the effect remained significant, $F(1, 17) = 7.941$, $p = .012$, p. $\eta^2 = 0.318$.

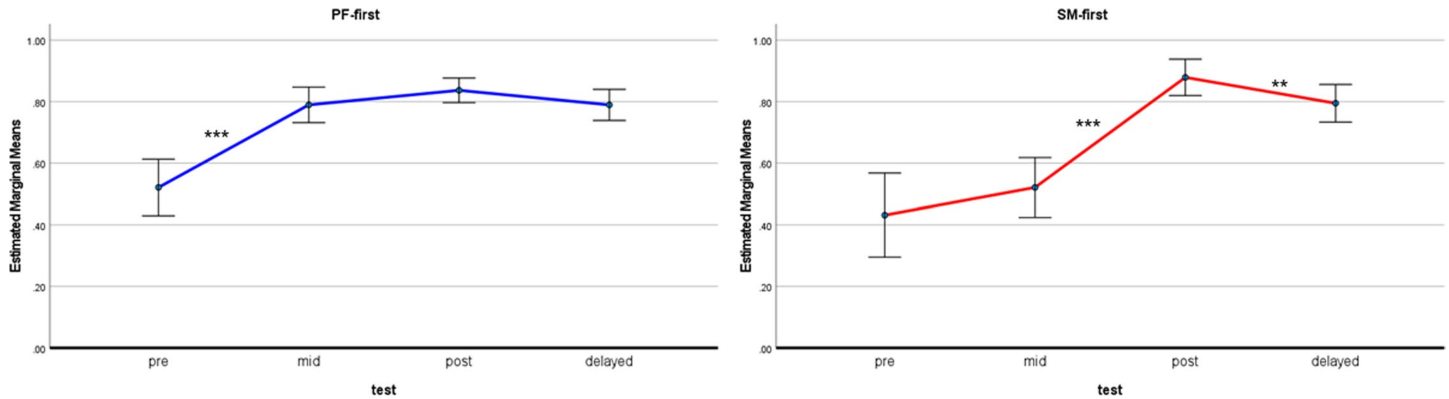

Fig. 8 Within group comparisons by condition after adjusting for pre-test ability, Experiment 1. Estimated marginal means of test scores by test-time with within-group comparisons. Conditions received their first intervention between the pre-test and the mid-test, and the second intervention between mid-test and post-test. Error bars represent 95% confidence intervals. ***: $p < .001$, **: $p < .01$, *: $p < .05$ (adjusted by Bonferroni correction)

perceptual-fluency intervention into their studying routines if available (91%, $n=40$). An ANOVA showed no differences between conditions on this question ($F<1$).

In sum, the interventions were well-received by participants who felt that while the tests were difficult, they did learn to perceive surgical anatomy better because of the interventions. Most participants preferred the perceptual-fluency intervention to the sense-making intervention, and most participants would use a perceptual-fluency intervention if available.

Discussion

Experiment 2 revealed a sequencing effect, with slightly increased learning gains when the sense-making intervention was presented before the perceptual-fluency intervention, but this advantage was not durable. Furthermore, we found significant learning gains only after participants received the perceptual-fluency intervention. This finding suggests that the sense-making intervention used in this experiment did not significantly enhance learning of surgical anatomy recognition. Taken together, this suggests that the perceptual-fluency intervention was both necessary and sufficient for learning gains in perceptual knowledge.

General discussion

We reported on two experiments that examined the effects of perceptual-fluency and sense-making interventions on perceptual learning of surgical anatomy. Previous work had shown that perceptual-fluency interventions are not only very effective at improving perceptual knowledge, but also at facilitating learning gains in content knowledge that primarily involved conceptual reasoning (Kellman et al., 2010; Rau et al., 2017a; Rau & Wu, 2018). However, prior research did not address the reverse scenario—whether sense-making interventions also contribute to learning gains of primarily perceptual knowledge.

Therefore, our goal was to investigate RQ1 (whether the sense-making intervention contributes to learning of perceptual knowledge), which we addressed in the context of professional training, where participants have extensive experience with the targeted content but are learning to apply their knowledge in practical situations. Experiment 1 revealed that the sense-making intervention did not enhance participants' perceptual learning of surgical anatomy compared to the perceptual-fluency intervention alone. This finding is noteworthy given that participants in the SM + PF condition received an additional intervention and therefore spent more time with the learning materials compared to participants in the PF-only condition. The SM + PF condition provided the sense-making intervention before the perceptual-fluency intervention because prior research had shown that prior knowledge determines which order of sense-making and perceptual-fluency interventions is most effective for participants' learning. However, this work had been done in a context where participants had little relevant prior experience with the targeted content.

Experiment 2 sought to address this limitation by investigating RQ 2.2 (whether the order of sense-making and perceptual-fluency interventions affects learning of perceptual knowledge). Our results showed an advantage of receiving the sense-making intervention first, which yielded higher learning outcomes at the post-test. Results from the survey showed that participants in this condition perceived the tasks as more difficult, which suggests that they may have experienced desirable difficulties that resulted in better performance on the post-test. However, the advantage of receiving the sense-making intervention first was only temporary and did not persist at the delayed-test. Further, our examination of learning gains within conditions showed that both conditions experienced significant learning gains only immediately after receiving the perceptual-fluency intervention, and not after receiving the sense-making intervention. Taken together, the findings from Experiment 1 and 2 suggested that the sense-making intervention altogether did not improve perceptual learning of surgical anatomy. Given that Experiments 1 and 2 focused on different surgical topics and used different sense-making interventions, we believe this finding is robust in the context of surgical training.

Finally, both experiments investigated how sense-making and perceptual-fluency interventions are received by surgery trainees. The results suggest that the interventions would be well-received by surgery trainees. Surveys following both studies suggested that trainees generally enjoyed the interventions and would incorporate them into their current studying routine.

These findings extend the current literature in several important ways. First, in contrast to prior research on STEM content learning, our results suggest that a perceptual-fluency intervention alone is sufficient to support perceptual learning of surgical anatomy. We attribute the difference between our findings and prior research to the fact that the learning goals in the prior STEM studies focused on content knowledge that involved complex reasoning about domain-relevant concepts. In contrast, the learning goal in our studies focused on simple recognition of perceptual patterns. Thus, our results are consistent with the *alignment hypothesis* (Koedinger et al., 2012), which suggests that complex learning goals (e.g., conceptual content knowledge) can be supported by a combination of simple and complex interventions (e.g., combined sense-making and perceptual-fluency interventions), whereas simple learning goals (i.e., perceptual knowledge) should be supported by simple interventions (i.e., perceptual-fluency interventions). Therefore, our findings do not contradict the prior research by Rau and Kellman, because their findings on combining sense-making and perceptual-fluency interventions are in line with the alignment hypothesis's prediction for complex learning goals. Our findings extend this prior work by confirming the alignment hypothesis for simple perceptual learning goals. Thus, our results suggest that targets of instruction that are inherently perceptual can be taught using perceptual-fluency interventions alone.

Our focus on perceptual targets of instruction also expands prior research by focusing on realistic visual representations. We consider two aspects of realism: perceptual richness and conceptual richness. The existing literature on sense-making and perceptual-fluency interventions has focused on relatively abstract representations that are perceptually sparse but conceptually rich. For example, a visual representation of an atom, while it may convey rich information about atomic structure, is designed to be perceptually sparse so that it can highlight key conceptual features. In contrast, the visual representations in the present work, which involved the identification of anatomy using surgical images, are perceptually rich but conceptually sparse. For example, a picture of an operative field is completely realistic, without color-coded structures, clear demarcations, or conventions for guidance (and is often obscured by a blurry camera lens or blood). Accurate perception of realistic images that are perceptually rich is more difficult than perception of human-designed abstract representations that are perceptually sparse (Mayer, 2002). In other domains, the difficulty associated with realistic images has been a disadvantage (Scheiter et al., 2009). However, perhaps unsurprisingly given the highly specific nature of the desired learning outcome of our experiment, our results suggest that targets of instruction involving perception of perceptually rich, realistic images can be taught using realistic perceptual-fluency interventions alone. This is consistent with Goldstone and Son's findings that concrete graphics outperformed idealized graphics when tested on immediate performance within that same domain (as opposed to assessing transfer) (Goldstone & Son, 2005).

In sum, our work established a boundary condition for the effectiveness of the combination of sense-making and perceptual fluency interventions by showing that this combination is not universally effective. Specifically, our results suggest that when the learning goal is perceptually rich and conceptually simple, a sense-making intervention is not needed in addition to a perceptual-fluency intervention. At the same time, this highlights that it is important to understand both the perceptual and conceptual richness of a learning task to determine whether sense-making and perceptual-fluency interventions should be combined. Hence, future work is needed to further locate this boundary on the spectrum of tasks that involve varying levels of perceptual and conceptual richness.

Finally, our results have practical implications. In many diverse fields, including aviation (Kellman & Massey, 2013), music (Zatorre, 1979), electrical system design (Egan & Schwartz, 1979), language fluency (Strange & Jenkins, 1978), and games like chess and GO (Chase & Simon, 1973; Reitman, 1976), perceptual knowledge is important for successful performance. Yet, instruction rarely utilizes perceptual-fluency interventions (Kellman & Garrigan, 2009), suggesting opportunities for more effective instruction. This is true, for example, in our specific context of surgical education. Despite the acknowledged importance of sensory semiosis (i.e., the perception of visual and haptic cues in the operating room—a perceptually rich but conceptually sparse task), surgical trainees learn anatomy primarily via sense-making methods (Cadieux et al., 2021; Cope et al., 2015a, b). Yet, our experiments showed no evidence of perceptual learning gains from sense-making interventions alone. Thus, the addition of perceptual-fluency interventions to surgical education could greatly improve the ability of trainees to effectively recognize anatomic structures. Given that errors in perception have been linked to devastating surgical complications such as bile duct injury (Way et al., 2003), improving perceptual knowledge of surgeons may have important benefits for patients. In principle, any medical field in which perception of rich, realistic images is an important target of instruction could benefit from perceptual-fluency interventions. In this way, our work likely has broad implications for medical fields such as dermatology, pathology, and radiology, and other non-medical fields that rely on perceptual expertise.

Limitations

Our results should be interpreted in relation to the following limitations. First, our experiments were performed with trainees at a single center. While it is unlikely that our participants' cognitive and perceptual abilities were systematically different from trainees at other medical centers, a multi-center study would provide stronger evidence for generalizability.

Second, there was some overlap of participants between Experiments 1 and 2. However, our two experiments focused on qualitatively different anatomic structures. Additionally, there was a break of several months duration between the two experiments. However, some familiarity with the protocols of the experiment may have been retained and impacted the results. Additionally, the experiments differed in other ways. Experiment 1 used a video-based sense-making intervention, while Experiment 2 used a text-based. This limits comparability between experiments, but may enhance generalizability given the consistency of the findings. Our experiments also did not contain checks of processing for the sense-making intervention, which would have confirmed that participants fully engaged in those interventions even though they were tested in perceptual fluency metrics.

Third, for experimental purposes, we have treated sense making and perceptual fluency as fully distinct entities and assessed their impact on a narrow, perceptual, aspect of surgical performance. Most tasks fall somewhere on a continuum between sense making and perceptual fluency and there is at least some marginal benefit from both types of learning. Further study will delineate these relative benefits with finer granularity.

Fourth, operating is a dynamic task, while the interventions and tests used in this study used static images. While this would seem to be a misalignment, we elected to use static images in our study for several reasons. First, the operating field is quite static. While certain structures can move despite the paralytics used with anesthesia (blood vessels, occasional bowel peristalsis), most do not. Movement, when it occurs, provides additional information about the underlying anatomy. Thus, the static images used in this study are more difficult to identify than dynamic clips would have been. This suggests that any gains seen in static image recognition would still apply to actual surgery. This is further supported by the dose-dependent relationship between operative experience and static image anatomy identification seen in our pre-test data (see Sect. 3.1.4 and 4.1.4). Increased surgical experience correlated positively with the ability to identify static anatomic structures. Finally, static images were technically easier to program anatomic image selection into than a looping video.

Fifth, while the use of two different sense-making interventions in Experiments 1 and 2 suggests that our findings are robust, other explanations are possible. Prior knowledge may have played a significant role as well. All trainees in our experiments had at least some prior knowledge of anatomy due to their medical training. While medical training tends to deliver this knowledge using non-realistic representations of anatomy (cadavers, illustrations), it may have adequately covered the conceptual knowledge that was conveyed by the sense-making intervention. These non-realistic representations of anatomy have several affordances that are well-matched to sense-making competencies: cadavers demonstrate anatomy in 3 dimensions and contain all anatomic structures of study in their native positions, while illustrations can be exaggerated and color-coded to highlight different features of anatomy. However, they are poor tools for teaching perceptual fluency—real anatomy is not color-coded, and even cadaver tissue is discolored, bloodless, and desiccated, appearing nothing like tissue in the operating room. Thus, the average medical student has much more preparation in anatomic sense-making than perceptual fluency. Future work with a study population without prior knowledge could be done to control for this confounding factor. Nevertheless, this highlights the importance of aligning the instructional methods with not only the learning goal, but also with the characteristics of the learner. In the context of professional training, where perceptual learning targets are common, students are likely to have prior conceptual knowledge.

Conclusion

To our knowledge, the present experiments are the first to systematically examine the effects of sense-making and perceptual-fluency interventions on perceptual learning in the context of professional training. Our results showed that our sense-making intervention did not enhance perceptual learning, whereas the perceptual-fluency intervention did. This suggests that perceptual targets of instruction require perceptual-fluency interventions and are not additionally aided by sense-making interventions. These findings are in line with Koedinger and colleagues' (2012) *alignment hypothesis* and establish a boundary condition of prior work suggesting that sense-making and perceptual-fluency interventions should be combined. Our work yields a pathway for future research to further locate this boundary along the spectrum of conceptually and perceptually rich learning goals. Further, our work has significant practical implications for perceptually rich learning contexts, such as surgical training, where common interventions focus on sense making rather than perceptual fluency, which presents an unfortunately common misalignment between the target of instruction and the instructional intervention.

Funding Christopher C. Stahl was supported by the National Cancer Institute of the National Institutes of Health under Award Number T32 CA090217. This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Declarations

Conflict of interest We have no known conflict of interest to disclose.

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, *33*(2), 131–152. https://doi.org/10.1016/S0360-1315(99)00029-9.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001.

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the Effects of different multiple Representational Systems in Learning Primary Mathematics. *Journal of the Learning Sciences*, *11*(1), 25–61. https://doi.org/10.1207/S15327809JLS1101_2.

Amiri, E., Sha, P., & Palmer, E. M. (2020). Training novices to discriminate retinal diseases using perceptual learning. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64*(1), 1456–1460. https://doi.org/10.1177/1071181320641347

Bodemer, D., & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior*, *22*(1), 27–42. https://doi.org/10.1016/j.chb.2005.01.005.

Cadieux, D. C., Mishra, A., & Goldszmidt, M. A. (2021). Before the scalpel: Surgical resident's preoperative preparatory strategies. *Medical Education*. https://doi.org/10.1111/medu.14449

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and Novices*. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology*, *20*(3), 396–413. https://doi.org/10.1080/09541440701728623.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.

Cope, A. C., Bezemer, J., Kneebone, R., & Lingard, L. (2015a). You see?" Teaching and learning how to interpret visual cues during surgery. Medical Education, 49(11), 1103–1116. https://doi.org/10.1111/medu.12780.

Cope, A. C. M., Mavroveli, S., Bezemer, J., Hanna, G. B. F., & Kneebone, R. F. (2015b). Making meaning from sensory cues: A qualitative investigation of Postgraduate Learning in the operating room. Academic Medicine, 90(8), 1125–1131. https://doi.org/10.1097/ACM.0000000000000740.

Dreyfus, H. L., & Dreyfus, S. E. (1986). Five steps from novice to expert. *Mind over machine.* NY: The Free Press.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, *7*(2), 149–158. https://doi.org/10.3758/BF03197595.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, *12*(1), 13–18. https://doi.org/10.1111/1467-8721.01213.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. https://doi.org/10.1016/S0364-0213(83)80009-3.

Gibson, E. J. (2000). Perceptual learning in development: Some Basic Concepts. *Ecological Psychology*, *12*(4), 295–302. https://doi.org/10.1207/S15326969ECO1204_04.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four … or is it two? *Memory (Hove, England)*, *12*(6), 732–747. https://doi.org/10.1080/09658210344000530.

Goldstone, R. L., Schyns, P. G., & Medin, D. L. (1997). Learning to bridge between perception and cognition. *Psychology of learning and motivation* (pp. 1–14). Academic Press.

Goldstone, R. L., & Son, J. Y. (2005). The transfer of Scientific Principles using concrete and idealized simulations. *Journal of the Learning Sciences*, *14*(1), 69–110. https://doi.org/10.1207/s15327809jls1401_4.

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, *6*(2), 53–84. https://doi.org/10.1016/j.plrev.2008.12.001.

Kellman, P. J., & Krasne, S. (2018). Accelerating expertise: Perceptual and adaptive learning technology in medical learning. *Medical Teacher*, *40*(8), 797–802. https://doi.org/10.1080/0142159X.2018.1484897.

Kellman, P. J., & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. *Psychology of learning and motivation* (pp. 117–165). Elsevier.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science*, *2*(2), 285–305. https://doi.org/10.1111/j.1756-8765.2009.01053.x.

Kellman, P. J., Massey, C., Roth, Z., Burke, T., Zucker, J., Saw, A., Aguero, K. E., & Wise, J. A. (2008). Perceptual learning and the technology of expertise: Studies in fraction learning and algebra. *Pragmatics & Cognition*, *16*(2), 356–405. https://doi.org/10.1075/pc.16.2.07kel.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science*, *36*(5), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x.

Kozma, R., & Russell, J. (2005). Students becoming chemists: Developing representational competence. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 121–145). Springer.

Krasne, S., Hillman, J. D., Kellman, P. J., & Drake, T. A. (2013). Applying perceptual and adaptive learning techniques for teaching introductory histopathology. *Journal of Pathology Informatics*. https://doi.org/10.4103/2153-3539.123991. *4*.

Krasne, S., Stevens, C. D., Kellman, P. J., & Niemann, J. T. (2020). Mastering electrocardiogram interpretation skills through a perceptual and adaptive learning module. *AEM Education and Training, 5*, e10454

Lowe, R. K. (1993). Constructing a mental representation from an abstract technical diagram. *Learning and Instruction*, *3*, 157–179. https://doi.org/10.1016/0959-4752(93)90002-H.

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation*, *41*, 85–139. https://doi.org/10.1016/S0079-7421(02)80005-6.

Nathan, M. J., Koedinger, K. R., & Alibali, M. W. (2001). Expert blind spot: When content knowledge eclipses pedagogical content knowledge. *Annual Meeting of the American Educational Research Association, 6*, 1.

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29*(4), 717–761. https://doi.org/10.1007/s10648-016-9365-3.

Rau, M. A. (2018). Sequencing support for sense making and perceptual induction of connections among multiple visual representations. *Journal of Educational Psychology*, *110*(6), 811–833. https://doi.org/10.1037/edu0000229.

Rau, M. A., Aleven, V., & Rummel, N. (2017a). Supporting students in making sense of connections and in becoming perceptually fluent in making connections among multiple graphical representations. *Journal of Educational Psychology*, *109*(3), 355–373. https://doi.org/10.1037/edu0000145.

Rau, M. A., Aleven, V., & Rummel, N. (2017b). Making connections among multiple graphical representations of fractions: Sense-making competencies enhance perceptual fluency, but not vice versa. *Instructional Science*, *45*(3), 331–357. https://doi.org/10.1007/s11251-017-9403-7.

Rau, M. A., Sen, A., & Zhu, X. (2019). Using machine learning to overcome the expert blind spot for perceptual fluency trainings. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial intelligence in education* (pp. 406–418). Springer International Publishing.

Rau, M. A., & Wu, S. P. W. (2018). Combining instructional activities for sense-making processes and perceptual-induction processes involved in connection-making among multiple visual representations. *Cognition and Instruction*, *36*(4), 361–395. https://doi.org/10.1080/07370008.2018.1494179.

Reitman, J. S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology*, *8*(3), 336–356. https://doi.org/10.1016/0010-0285(76)90011-6.

Richman, H. B., Gobet, F., Staszewski, J. J., & Simon, H. A. (1996). Perceptual and memory processes in the acquisition of expert performance: The EPAM model. *The road to excellence: The acquisition of* *expert performance in the arts and sciences, sports, and games* (pp. 167–187). Lawrence Erlbaum Associates, Inc.

Romito, B. T., Krasne, S., Kellman, P. J., & Dhillon, A. (2016). The impact of a perceptual and adaptive learning module on transoesophageal echocardiography interpretation by anaesthesiology residents. *British Journal of Anaesthesia*, *117*(4), 477–481. https://doi.org/10.1093/bja/aew295.

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction*, *19*(6), 481–494. https://doi.org/10.1016/j.learninstruc.2008.08.001.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 72–103). Cambridge University Press.

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). *At a loss from words: Verbal overshadowing of perceptual memories* (pp. 291–340). Academic Press.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*(2), 227–237. https://doi.org/10.1016/S0959-4752(02)00022-1.

Strange, W., & Jenkins, J. J. (1978). Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick (Eds.), *Perception and experience* (pp. 125–169). Springer.

Way, L. W., Stewart, L., Gantert, W., Liu, K., Lee, C. M., Whang, K., & Hunter, J. G. (2003). Causes and prevention of laparoscopic bile duct injuries: Analysis of 252 cases from a human factors and cognitive psychology perspective. *Annals of Surgery*, *237*(4), 460–469. https://doi.org/10.1097/01.SLA.0000060680.92690.E9.

Wylie, R., Koedinger, K. R., & Mitamura, T. (2009). Is self-explanation always better? The effects of adding self-explanation prompts to an english grammar tutor. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 1300–1305.

Zatorre, R. J. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics*, *26*(5), 384–395.