European Journal of Psychology of Education
https://doi.org/10.1007/s10212-022-00616-x

# How are feelings of difficulty and familiarity linked to learning behaviors and gains in a complex science learning task?

Yingbin Zhang[1] · Luc Paquette[1] · Ryan S. Baker[2] · Nigel Bosch[3] · Jaclyn Ocumpaugh[4] · Gautam Biswas[5]

Received: 22 October 2021 / Revised: 4 March 2022 / Accepted: 27 March 2022


## Abstract

The feelings of difficulty and familiarity (FOD and FOF) are two types of metacognitive experiences. Both may influence student engagement and the application of metacognitive strategies, but these relationships are not well understood, in part because many studies have relied on self-report measures of behaviors that may not accurately reflect students' actual behaviors. In this study, FOD and FOF were related to objective measures of off-task behaviors and metacognitive strategies. These measures were extracted from 88 sixth graders' action logs within a computer-based learning environment known as Betty's Brain. Pre- and post-tests were administered to assess learning. Results reveal that high-FOD students showed more off-task behaviors and fewer strategic behaviors than low-FOD students, particularly when this difference was measured in terms of the frequency (as opposed to proportion) of strategic behaviors. FOF was not associated with off-task behaviors and metacognitive strategies but emerged as a moderator in the relationship between FOD and learning gains. Low-FOD students learned more than high-FOD students in the low-FOF group, but such a difference was not found in the high-FOF group.

**Keywords** Feeling of difficulty · Feeling of familiarity · Metacognitive experience · Metacognitive strategy · Off-task behavior

✉ Yingbin Zhang
yingbin2@illinois.edu

[1] Department of Curriculum & Instruction, University of Illinois at Urbana-Champaign, 1310 S Sixth Street, Champaign, IL 61820, USA

[2] Graduate School of Education, University of Pennsylvania, 3700 Walnut St, Philadelphia, PA 19104, USA

[3] School of Information Sciences and Department of Educational Psychology, University of Illinois at Urbana-Champaign, 501 E. Daniel St, IL 61820 Champaign, USA

[4] Penn Center for Learning Analytics, University of Pennsylvania, 3700 Walnut St, Philadelphia, PA 19104, USA

[5] Department of Computer Science/ISIS, Vanderbilt University, 1025 16th Ave South, Nashville, TN 37212, USA

Published online: 11 April 2022

## Introduction

Metacognitive strategies are important for effective learning in complex domains (Pintrich et al., 1993), specifically those requiring complex monitoring and control activities (Azevedo & Aleven, 2013). As researchers have worked to understand the relationship between metacognitive strategies and learning (Ohtani & Hisasaka, 2018; Veenman et al., 1997), they have also sought to better understand metacognitive experiences that influence the use of metacognitive strategies (Flavell, 1979). In particular, researchers have identified the importance of two metacognitive experiences: the feeling of difficulty (FOD) and the feeling of familiarity (FOF) (Efklides et al., 1999; Garcia-Marques & Mackie, 2001; Malmberg et al., 2013; Whittlesea & Williams, 1998).

Studies have found that FOD and FOF may be associated with how students approach learning tasks, including their engagement and their use of metacognitive strategies (Beckmann & Goode, 2014; Patall et al., 2018; Rellinger et al., 1995; Soemer et al., 2019). Most of these studies have assessed engagement and metacognitive strategies use via self-report instruments. Self-report tools are appropriate for measuring FOD and FOF because they are individuals' subjective feelings. Students are aware of their feelings and use them to guide the regulation of learning (Efklides, 2006, 2009). However, self-report measures of behaviors are more problematic, with previous research showing that self-report behaviors may be poorly correlated with actual behaviors (Azevedo, 2015; Craig et al., 2020). Objective measures, such as those extracted from action logs in computer-based learning environments, may be more precise in capturing task-specific behaviors and strategies (Rovers et al., 2019).

Moreover, previous studies have examined the links among FOD, FOF, and learning behaviors during such tasks as mathematics problem-solving (Efklides et al., 1999), information searching (Liu et al., 2012), and informative text reading (Soemer et al., 2019). However, no study has investigated these relationships in a complex science learning task.

The current study attempts to address these gaps by exploring the relationship between FOD, FOF, learning, and objective measures of strategic and off-task behaviors in learning about human thermoregulation within a computer-based learning environment, Betty's Brain (Biswas et al., 2016). Specifically, we compare concrete behavioral measures, as extracted from students' action logs in Betty's Brain, to awareness of subjective states, including FOD and FOF. The rest of the introduction discusses the theoretical distinctions and associations among metacognitive experiences, strategies, and knowledge; reviews studies about FOF and FOD; and presents the research questions that underpin the current research.

### Metacognitive knowledge, strategies, and experience

Metacognition includes three facets: metacognitive knowledge, metacognitive strategies, and metacognitive experiences (Efklides, 2008). Metacognitive knowledge primarily refers to declarative knowledge and beliefs about self, cognition, and past task-processing experiences. Metacognitive strategies refer to the procedural knowledge of how to regulate cognition. Examples are monitoring, control, and self-evaluation (Dent & Koenka, 2016). Finally, metacognitive experiences refer to experiences that we are consciously aware of while engaging in tasks. It takes the form of metacognitive feelings, metacognitive judgments, and online task-specific knowledge. Metacognitive feelings, such as FOD and FOF, are the products of non-analytic and nonconscious inferential processes (Efklides, 2006). They reflect how learners feel they are doing while engaging in the task. Online task-specific knowledge is the knowledge of the task that we are aware of in task processing. It is analytic. Metacognitive judgments are judgments about learning, the demand of a task, the task solution, etc., which can be analytic or non-analytic.

Nelson and Narens' (1990) two-level metacognitive system connects metacognitive knowledge, strategies, and experiences in a unified framework. This framework distinguishes between object-level processing, which describes cognition about the external world, and meta-level processing, or cognition about cognition. The meta-level receives information about the object-level via a monitoring function and sends orders to regulate the object-level activities via a control function. The monitoring process uses metacognitive knowledge and triggers metacognitive experiences (Efklides, 2006). For example, if the meta-level detects an interruption in the object-level activities, learners may experience a feeling of difficulty. Learners then draw upon their metacognitive experiences and knowledge to activate the meta-level control function (Efklides, 2008), which manifests as the application of cognitive skills and metacognitive strategies (Efklides, 2006).

Recently, Efklides' (2011) metacognitive and affective model of self-regulated learning (MASRL) provides an account for the impact of metacognitive experiences on the application of metacognitive strategies. In this model, cognitive, metacognitive, motivational, and affective factors interact during learning processes. During task processing, metacognitive experiences contribute to motivation and affect, which, in turn, modify the evaluation of prior decisions and drive behavior regulation. Metacognitive experiences may also influence the metacognitive strategy behaviors directly. Overall, as metacognitive experiences, FOD and FOF have meaningful theoretical relationships with learning behaviors.

## Feeling of difficulty (FOD)

FOD describes the subjective feeling that arises when a learner interacts with a task that they are unfamiliar with, that demands heavy working memory, or that involves events that are discrepant with their knowledge structure (Touroutoglou & Efklides, 2010). As such, it is a product of the interaction among individuals' ability, self-concept, and the task (Efklides & Tsiora, 2002; Efklides et al., 1998).

FOD is associated with a range of phenomena in cognition around learning. Studies have used FOD as a subjective measure of cognitive load and found it related to mental effort appraisals, another subjective measure of cognitive load (Ayres, 2006; Schmeck et al., 2015). The latter refers to the amount of mental effort investment that learners reported in completing a task. FOD may drive effort investment, but if FOD is extremely high, learners may also avoid putting effort into the task (van Gog & Paas, 2008).

Learners who report a higher FOD toward a task tend to exhibit lower situational interest (Fulmer & Tulis, 2013), more negative affective experiences (Fulmer & Tulis, 2013), lower perceived competency (Tulis & Fulmer, 2013), and less emotional and behavioral engagement with the task (Patall et al., 2018). By contrast, other studies have found that FOD is positively associated with strategy use during recall tasks (Rellinger et al., 1995) and mathematics problem-solving (Efklides et al., 1999). The reason may be that FOD promotes effortful and analytic information processing (Alter et al., 2007) and triggers metacognitive control functions that result in the learner invoking cognitive and metacognitive strategies (Efklides, 2011).

Thus, higher FOD has been related to more disengagement in some studies (e.g., Fulmer & Tulis, 2013; Patall et al., 2018) but more application of learning strategies in others (e.g., Efklides et al., 1999; Rellinger et al., 1995). This raises the question of whether the relationship between disengagement and strategy use is necessarily negative. Thus far, no study has investigated the associations among FOD, disengagement, and strategy use together. Moreover, the measurements of behaviors in most prior studies were self-reported (e.g., Patall et al., 2018), and they may be weakly related to actual disengagement and strategy use (Craig et al., 2020). It is, therefore, worth examining the associations between FOD and behavioral measures in learning.

## Feeling of familiarity (FOF)

FOF arises when the current task is closely tied to previous experiences or when participants attribute the fluency on the current task to prior experiences (Efklides, 2006; Whittlesea, 1993). The current task can be an exact repetition or semantically related to prior tasks. FOF is not based on knowledge and distinct from the feeling of knowing (Efklides, 2009; Kinoshita, 1997), which is the probability reported by individuals that they cannot retrieve a piece of information now but will be able to do it later on.

Educators have suggested that learning material should be related to students' prior experiences (Merrill, 2002; Rivet & Krajcik, 2008), and that FOF may induce situational interest in a given task (Alexander et al., 1994; Soppe et al., 2005). Empirically, FOF was associated with less mind wandering during reading activities (Soemer et al., 2019). In addition, English as a second language learners showed more behavioral and cognitive engagement on familiar oral narrative tasks than on unfamiliar tasks (Qiu & Lo, 2016).

Still, prior experience is not always beneficial (Reder et al., 2007), and familiarity can cause interference. For example, Beckmann and Goode (2014) found that students in a semantically familiar context acquired less knowledge about abstract scientific principles and performed worse in a subsequent task than students in a less familiar context. Further analyses found that the former held more a priori assumptions about the scientific principles than the latter, although they had no difference in prior knowledge. Students with more a priori assumptions tended to use less systematical exploration strategy during learning. The researchers concluded that students may form many a priori assumptions in a semantically familiar context but do not systematically test these assumptions during learning.

Likewise, Garcia-Marques and Mackie (2001) discussed what they call the familiarity-stereotype effect, which refers to the phenomenon that FOF promotes non-analytic information processing, while unfamiliarity induces analytic processing. This finding aligns with research showing that FOF increases the feeling of knowing (Reder & Ritter, 1992; Schwartz & Metcalfe, 1992), and when learners think they know something, they put less effort into studying it (Metcalfe, 2009). Indeed, Sockalingam and Schmidt's (2013) work on problem-based learning found that unfamiliar problems stimulated the most questioning and reasoning behaviors.

To summarize, higher FOF has been related to less behavioral disengagement in some studies (Qiu & Lo, 2016; Soemer et al., 2019) but less analytic cognitive processing and strategy use in the others (Garcia-Marques & Mackie, 2001; Sockalingam & Schmidt, 2013). Again, these findings raise the question of whether disengagement is negatively related to strategy use. No study has investigated the associations among FOF, disengagement, and strategy use together. Moreover, the role of FOF in learning complex scientific phenomena is not yet understood (Beckmann & Goode, 2014).

## Current research

Self-report measures of disengagement and metacognitive strategies have been related to FOD and FOF (Patall et al., 2018; Qiu & Lo, 2016). It is not clear that the same patterns will hold when behavioral measures of disengagement and metacognitive strategies are utilized, given the differences between self-report and behavioral measures of these constructs (Craig et al., 2020). These relationships have also not yet been studied in the context of learning complex scientific phenomena. This study addresses these gaps in the context of learning human thermoregulation within a computer-based environment, Betty's Brain (Biswas et al., 2016). Students' action logs in Betty's Brain can be used to extract behavioral measures of disengagement and metacognitive strategies. Therefore, the first two research questions are: what are the relationships among FOD, FOF, and off-task behaviors as an indicator of disengagement (RQ1) and the relationships among FOD, FOF, and metacognitive strategy behaviors (RQ2)?

RQ2 explores the relationships between FOD, FOF, and two different measures of metacognition, namely the frequency of the various types of behaviors (RQ2.1) and the proportion of them that can be classified as metacognitive (RQ2.2). We consider both frequencies and proportions because they may have different associations with FOD and FOF. For example, previous research (Alter et al., 2007; Garcia-Marques & Mackie, 2001) has found that high FOD and low FOF may be related to effortful and analytic cognitive processing. Meanwhile, Evans and Stanovich (2013) suggest that analytic processing is slower than non-analytic cognitive processing. If students with high FOD and low FOF have slower cognitive processing, we might expect them to execute fewer actions and metacognitive strategies (e.g. a lower frequency).

Likewise, we consider the proportion of each behavior type that reflects metacognition. Betty's Brain includes six behavioral categories (e.g., reading and taking quizzes), five of which were subclassified as metacognitive and not metacognitive. For example, reading behaviors were classified as metacognitive when they were coherent with quiz results preceding them because the coherence suggested that the student might intentionally seek relevant information to improve their understanding based on the quiz result (see the "Metacognitive strategies" section). This allowed us to calculate the proportion of each behavioral category (e.g., reading) that was metacognitive for each student.

Analytic cognitive processing allows learners to identify discrepancies between their progress and their learning goal and may trigger the application of metacognitive strategies to regulate cognition (Efklides, 2011). Such processing might lead to lower frequency of strategy use but a relatively high proportion of strategy use in Betty's Brain, given that students' time on Betty's Brain was not strictly limited. Although students with high FOD may conduct metacognitive strategies less frequently than those with low FOD due to slow cognitive processing, the two groups' proportions of metacognitive strategy behaviors may be close. Similarly, the low-FOF group's proportion of metacognitive strategy behaviors may be close to the high-FOF group's. Thus, we may expect no difference in the proportion of metacognitive strategy behaviors between the high- and low-FOD groups and between the high- and low-FOF groups, or that the differences are larger in the frequency than in the proportion.

More off-task behaviors and fewer metacognitive strategy behaviors may impair learning, and empirical studies have found associations among FOD, FOF, and task performance in mathematics (Efklides et al., 1998; Yang et al., 2019). Thus, the third

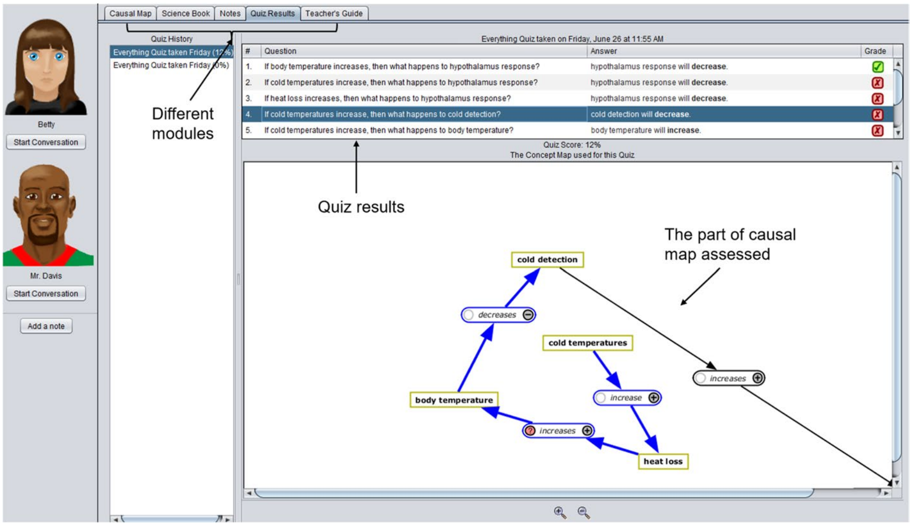


**Fig. 1** Screenshot of viewing quiz results in Betty's Brain. The upper right shows the quiz questions, answers, and grades. The fourth question, which was answered incorrectly, was selected, and the concepts and links that Betty used to answer this question were highlighted

research question is (RQ3): what are the relationships among FOD, FOF, and learning gains (operationalized as the change between pre- and post-test scores) in learning about complex scientific phenomena?

It is worth noting that FOD and FOF are not mutually exclusive. For example, FOD may increase when the task is familiar but demands a heavy working memory load (Touroutoglou & Efklides, 2010). Consequently, FOD and FOF may interactively impact the learning process. Thus, we test the interaction between FOD and FOF on learning behaviors (in RQ1 and RQ2) and learning gains (in RQ3).

## Methods

### Betty's Brain

Betty's Brain is a computer-based learning environment where students learn about scientific phenomena such as climate change and thermoregulation (Biswas et al., 2016). It uses a learning-by-teaching approach: students learn science by teaching a virtual agent named Betty. Teaching consists of building a causal map about the scientific phenomenon in which causal relationships are represented by directed causal links among concepts (see Fig. 1). To build this map, students have access to hypermedia resource pages on relevant scientific concepts (*Science Book* in Fig. 1). Students can evaluate their causal map by having Betty answer questions. By looking at Betty's correct and incorrect answers (see Fig. 1), students can identify problems in their causal map (e.g., incorrect links). They improve their understanding and fix the problems in the map by reading relevant resource pages and revising the incorrect links.

There is a virtual pedagogical agent, Mr. Davis, who provides students with support for building the map. If the student is not making progress (e.g., "quiz score has not improved in the students' last five attempts at updating their map"), Mr. Davis may prompt students to read resource pages containing information that could improve the causal map (e.g., "You should go and read the page on Heat Loss").

In this study, students learned about human thermoregulation—the human body's processes when exposed to cold temperatures. Thermoregulation is a complex physiological phenomenon, and a basic understanding of it is essential for studying biology and biology-related disciplines (Tansey & Johnson, 2015). The text about this topic in Betty's Brain had 15 pages and 1974 words, covering 13 scientific concepts and 15 causal links.

## Participants and procedures

During the 2018–2019 school year, data were collected from 88 sixth graders from four classrooms in a southern US urban public school. The school served around 700 students in grades 5–8 and reported a student population that was 60% White, 25% Black, 9% Asian, and 5% Hispanic. Around 8% were enrolled in the free and reduced-price lunch program. No demographic data were collected from individual students due to privacy issues.

The study lasted 6 school days, during which time students worked individually on laptops (but were often seated at tables of 4–5 students). On day 1, students spent 30–45 min completing a paper-based pre-test. On day 2, they received a 30-min training about how to use Betty's Brain. In the next 3 days, they spent around 30 min per day teaching Betty about thermoregulation by constructing the causal map. On the final day, students first completed a metacognitive experience questionnaire and a science anxiety questionnaire (not included in this study), and then, a post-test. The metacognitive experience questionnaire contained items that asked about students' FOD and FOF toward the Betty's Brain unit on thermoregulation (current unit) and a unit on climate change that was completed 1 month before the current study. As it had been 1 month after learning about climate change, FOD and FOF toward this unit might not be as accurate as students' FOD and FOF toward the human thermoregulation unit. Thus, we did not analyze the associations among FOD, FOF, learning behaviors, and learning gains in the climate change unit.

## Measures

### Metacognitive experiences

Based on Efklides' (2002) work, two items were used to measure FOD and FOF. Research has shown that the one-item measures of metacognitive experiences were consistent over the course of solving a task in a computer-based learning environment (Dindar et al., 2020). The FOD item asked students to rate "how hard was the Betty's Brain's unit on thermoregulation."[1] The choices included *1 = very easy*, *2 = easy*, *3 = neither*, *4 = difficult*, and *5 = very difficult*. The FOF item asked students to rate "how familiar were you with the science concepts in the thermoregulation unit." The choices included *1 = very unfamiliar*,

[1] Readers unfamiliar with Betty's Brain may think the question is asking the difficulty of the concept of thermoregulation rather than the difficulty of learning thermoregulation. However, a Betty's Brain unit mainly refers to the task of building a causal map for the unit rather than the unit's resource book. We believe that students considered the Betty's Brain's unit on thermoregulation in terms of this perspective because they had received training on Betty's Brain and used the system for several days.

*2 = unfamiliar*, *3 = neither*, *4 = familiar*, and *5 = very familiar*. It is worth highlighting that metacognitive experiences are subjective experiences that students are aware of (Efklides, 2006, 2009). Thus, we argue that it is reasonable to assess students' metacognitive experiences via self-report questionnaires.

Past research has primarily used four, five, seven, or ten response categories for FOD and FOF items (e.g., Efklides, 2002; Efklides et al., 1999; Fulmer & Tulis, 2013; Liu et al., 2012; Soemer et al., 2019; Touroutoglou & Efklides, 2010; Tulis & Fulmer, 2013). However, seven- and ten-point scales were mainly used in samples of older students such as undergraduates (Liu et al., 2012; Touroutoglou & Efklides, 2010). Borgers et al. (2004) suggested that fewer response categories may be better when the participants are children. Offering four or five categories rather than more may reduce the cognitive load when participants respond to the item (Dillman et al., 2014). Thus, four or five categories may be appropriate for sixth graders. We chose a five-point scale rather than a four-point scale to allow neutral responses if students felt that the task was neither difficult nor easy or neither familiar nor unfamiliar.

Similar items measured FOD and FOF toward the climate change unit. Although we did not analyze the data in the climate change unit, the differences in FOD and FOF between the climate change and thermoregulation units may provide evidence about the validity of FOD and FOF measurements. Both units' Flesch–Kincaid reading grade levels were 8.0, but students felt that the climate change unit was easier than the thermoregulation unit ($t = 3.38$, $df = 87$, Cohen's $d = 0.36$, $p = 0.001$). They felt more familiar with the climate change unit than the thermoregulation unit ($t = 11.71$, $df = 87$, Cohen's $d = 1.26$, $p < 0.001$). The differences match our expectations as climate change is a relatively popular science topic. Young people (aged 12–25) are likely to know something about climate change and have interests and concerns about it (Corner et al., 2015). The concepts of climate change (e.g., carbon dioxide and greenhouse effect) may be more common than the concepts of human thermoregulation (e.g., cold detection and hypothalamus response) in modern media. Thus, the differences in FOD and FOF between the two units may provide evidence that the FOD and FOF items measured what they intended to measure, i.e., construct validity.

### Metacognitive strategies

Metacognitive strategies were inferred using coherence analysis, a process that identifies coherent actions (Segedy et al., 2015). Coherent actions are two ordered actions where the first action generates information supporting the second action. For instance, in Fig. 1, the quiz results could inform students that the question about the causal relation between cold temperature and detection was answered incorrectly. After viewing these quiz results, if students read the resource pages that contains information about the relationship between the two concepts, the viewing and the reading actions are defined as coherent. The two actions need not be consecutive, but their time interval is restricted to avoid coincidental connections. Prior research in Betty's Brain found that students usually use information within 5 min of encountering the information (Segedy, 2014). The proportion of actions not supported by prior actions within 5 min was negatively related to map scores (the number of correct causal links minus the number of incorrect links), while the proportion of information that was used within 5 min was positively related to map scores and the change between pre-test and post-test scores (Segedy et al., 2015). Thus, we set the interval restriction for the current study to 5 min.

To execute coherent actions, students must monitor previous cognitive activities and the information generated by previous activities (the first action and its results, e.g., viewing quiz results generated the information that the link chain between cold temperature and detection was wrong). Then, they need to regulate the current behavior (the second action, e.g., reading resource pages about concepts involved in the link chain) based on the received information. Therefore, coherent actions may suggest the use of metacognitive strategies (Zhang et al., 2020; Segedy et al., 2015).

The Betty's Brain system logs all of the actions a student perform along with the context in which they were performed. This makes it possible to evaluate whether a series of actions represents a coherent strategy or a more random approach. Specifically, we used the log files to label five types of coherent actions:

1. *Coherent viewing*—viewing quiz results that were semantically consistent with subsequent actions. This consistency suggests that students were monitoring (e.g., Dent & Koenka, 2016) the information they had collected and deliberately using it in subsequent actions (Zhang et al., 2020).
2. *Coherent editing*—making edits (additions, deletions, or other changes) to the causal map in a way that is consistent with content students had just seen in the virtual textbook or in one of the quizzes. This suggests that students were using control strategies (e.g., Dent & Koenka, 2016) to edit the concept map in accordance with the information they had just acquired (Zhang et al., 2020).
3. *Coherent reading*—reading a resource page that reflects feedback (either the quiz results or the comments from Mr. Davis) that the student has just received (Biswas et al., 2016). This suggests the student intentionally (e.g., with control) sought relevant information to improve their understanding based on the quiz result (Zhang et al., 2020).
4. *Coherent marking*—marking (an annotation feature in the causal map) in a way that reflects recent quiz results. This suggests that the student understood what links on their map were correct or incorrect and annotated them accordingly. Coherent marking might represent constructive monitoring behaviors because the marking action translates quiz results into systematic checking of the causal maps (Zhang et al., 2020).
5. *Coherent feedback*—receiving feedback from Mr. Davis that is consistent with subsequent reading actions. In other words, the student received a prompt and then read the resource it suggested. This variable assessed whether students took advantage of the feedback from the learning environment and might reflect control behaviors.

Non-coherent actions also occur in Betty's Brain, and may be effective or ineffective. For example, consider a non-coherent reading action that occurs when a student reads a resource page that does not contain information useful for improving the causal map. Prior to this reading action, the causal links of this page had already been correctly added to the causal map. The reason that the student does this reading action may be (1) they fail to regulate behaviors based on the current progress and unnecessarily (perhaps haphazardly) read this page, or (2) they feel they do not understand the content of this page thoroughly and reread the page to reinforce their knowledge. The first reason suggests an ineffective strategy, but the second reason may suggest a more effective strategy, such as the monitoring (they were aware of insufficient understanding) and control strategy (they decided to reread the page). We could not distinguish these reasons based on the action log, so we did not analyze the efficacy of non-coherent actions.

For each kind of coherent action, we calculated two measures: the frequency (number) per minute and the proportion (i.e., the ratio of the number of coherent reading actions to the number of reading actions). We excluded actions that were too brief, including viewing (quiz results) actions shorter than 2 s and reading (resource pages) actions less than 10 s (no matter if these actions were coherent or not). These actions were unlikely to reflect the use of a metacognitive strategy (Segedy et al., 2015). For example, a very short time on viewing quiz results might indicate that students just skimmed the quiz results without analyzing the link's correctness.

### Off-task behavior

We defined off-task behaviors as cases where, within a period at least 5-min long, students did none of the following actions: (1) stay in a resource page for at least 30 s, (2) stay in a quiz result for at least 30 s, (3) edit a link, or (4) mark a link in Betty's Brain. Off-task behaviors also included cases where students stayed in a page or quiz result for more than 10 min, substantially longer than the time needed to read a page or interpret a quiz result carefully. We calculated the proportion of time on these behaviors as a measure of off-task behavior: the ratio of the time on these behaviors to the total time on Betty's Brain (including the off-task time). This metric was negatively related to the performance in Betty's Brain in prior research ($r = -0.46$; Segedy et al., 2015). The proportion of off-task time has also been used as an indicator of behavioral disengagement in other learning environments (Godwin et al., 2021; Henrie et al., 2015).

### Knowledge tests

The pre- and post-tests assessed students' knowledge of human thermoregulation and causal relationships. They were identical both in form and content. The test contained four causal reasoning items, eight multiple-choice items, and four short-answer items. A causal reasoning item consisted of a causal map, where abstract concepts (i.e., concepts were named X, Y, etc.) were connected unidirectionally, and a question about the map (e.g., If X decreases, what will happen to concept Y?). Each question had four choices: Y will (1) increase, (2) decrease, (3) not be affected, (4) depend on which causal relations are stronger. Multiple-choice items tested students' knowledge of the thermoregulation domain, and each had four choices. Students got one point if they answered a causal reasoning or multiple-choice item correctly. Short-answer items asked students to explain the human body's responses to cold temperatures based on their understanding of the causal relations among concepts in the domain. The correct answer to each item contained three to five successive causal links between a relevant set of concepts. A student got one point if their answers had one link the same or close to a link in the correct answer. The maximum possible test score was 27. Coefficient alpha was 0.60 and 0.80 for the pre- and post-tests, respectively. Coefficient alpha was a little lower in the pre-test because some items were answered correctly by only a few students, causing low variances in these item scores and low correlations between these item scores and the overall pre-test score. In the post-test, more students gave correct answers to these questions, indicating learning. Thus, these questions were kept in both the pre- and post-tests.

**Table 1** The distribution of FOD and FOF

| FOD/FOF | Very unfamiliar | Unfamiliar | Neither | Familiar | Very familiar |
|---|---|---|---|---|---|
| Very easy | 1 | 1 | 1 | 0 | 0 |
| Easy | 2 | 5 | 2 | 2 | 2 |
| Neither | 2 | 3 | 3 | 4 | 1 |
| Difficult | 15 | 16 | 3 | 2 | 0 |
| Very difficult | 14 | 4 | 3 | 1 | 1 |

## Analyses

The distributions of FOD and FOF were skewed (see Table 1). For FOD, there were few *very easy*, *easy*, or *neither* responses. For FOF, there were few *very familiar*, *familiar*, and *neither* responses. Treating these variables as continuous might therefore lead to biased results. Thus, we divided students' ratings for each variable into two categories: high-FOD (ratings of difficult and very difficult) and low-FOD (ratings of neither, easy, and very easy), as well as high-FOF (ratings of neither, familiar, and very familiar) and low-FOF (ratings of unfamiliar and very unfamiliar). This division was based on the meaning of the rating options and the consideration of as much size balance across groups as possible. Putting neither and easy or unfamiliar options together did not mean that the neither option had the same valence for students as the easy and unfamiliar options. Instead, it emphasized that the students who selected difficult options perceived the task as more difficult than those selecting neither and easy options, and that students who selected neither and familiar options perceived the task as more familiar than the students selecting unfamiliar options. The division generated four groups: high-FOD and high-FOF ($N=10$), high-FOD and low-FOF ($N=49$), low-FOD and high-FOF ($N=15$), as well as low-FOD and low-FOF ($N=14$).

The normality and homogeneity assumptions of most of the response variables, such as the frequency of coherent actions, were violated, and groups were unbalanced in size. Thus, we used non-parametric tests to examine the associations among FOD, FOF, and the other variables. Specifically, to examine the associations between FOD, FOF, and the proportion of off-task time (RQ1), we conducted a robust two-way ANOVA with 10% trimmed means (a robust measure of means that ignores the top and bottom 10% of data; Wilcox, 2011). The robust test is satisfactory even when normality and homoscedasticity assumptions are violated (Wilcox, 2011). The robust ANOVA produces a $\chi^2$-distributed test statistic, $Q$. The $p$ value is calculated by comparing $Q$ with an adjusted significance criterion (i.e., $\alpha$ value), so the degrees of freedom (*df*) are not reported (Mair & Wilcox, 2020). The robust ANOVA was implemented within the WRS2 package in R (Mair & Wilcox, 2020).

To examine the association between FOD, FOF, and coherent actions (RQ2.1 and RQ2.2), we conducted a robust two-way ANOVA with 10% trimmed means for each coherent metric. We used the Benjamini–Hochberg correction to control the false discovery rate (FDR) across multiple tests. This correction adjusts the $\alpha$ value rather than the $p$ value, so we only marked a result as statistically significant if its $p$ value was lower than the adjusted $\alpha$ value (the initial $\alpha$ was 0.05). We computed the partial omega squared statistics ($\omega_p^2$) since this metric is recommended for the comparison of effects across analyses with the same design (Lakens, 2013). It is an estimate of the proportion of response variables' variances accounted for by predictors. A rule of thumb for its interpretation is that 0.01 ~ 0.06, 0.06 ~ 0.14, and greater than 0.14 correspond to small, medium, and large effects, respectively (Field, 2013).

To examine the associations between FOD, FOF, and test scores (RQ3), we conducted a three-way, rank-based repeated ANOVA with the R package nparLD[2] (Noguchi et al., 2012). Test time (pre and post) was the within-subject factor, and FOD and FOF were between-subject factors yielding four groups (high/low FOD × high/low FOF). The repeated ANOVA produces an ANOVA-type statistic (ATS), which can be approximated by the *F* distribution with an infinity *df* in the denominator (Brunner & Puri, 2001). The nparLD package also generates the relative treatment effect, and an increase/decrease in this effect between test time represents an increase/decrease in the test scores.

## Results

Table 2 displays the means, standard deviations, and Kendall's Tau correlations among raw variables in this study. FOD was negatively correlated with FOF. Neither were correlated with the proportion of off-task time. FOD was negatively correlated with four coherence metrics, while FOF was only positively correlated with the proportion of coherent viewing. FOD and FOF were not correlated with pre-test scores, and we, therefore, did not use pre-test scores as a covariate when examining the relationships among FOD, FOF, off-task behaviors, and coherent metrics.

### Off-task behavior

Figure 2 displays different groups' proportions of off-task time. The robust ANOVA reveals a main effect of FOD ($Q = 6.55$, $\omega_p^2 = 0.05$, $p = 0.03$). High-FOD students showed a greater proportion of off-task time than low-FOD students. There was neither a main effect of FOF ($Q = 1.25$, $\omega_p^2 = 0.00$, $p = 0.28$) nor an interaction between FOD and FOF ($Q = 0.33$, $\omega_p^2 = 0.00$, $p = 0.55$).

### Coherent actions

Figures 3 and 4 display the frequency and proportion of various coherent actions. Overall, low-FOD groups showed greater frequency and proportion of coherent actions than high-FOD groups, but the differences between high-FOF and low-FOF groups were not consistent. The results of the robust ANOVA matched the figures, which were displayed in Table 3. A main effect of FOD was found on all measures of coherent actions ($Q = 4.60 \sim 11.63$, $p <$ adjusted $\alpha$), except the proportion of coherent edits ($Q = 1.11$, $p = 0.30$). There was no effect of FOF ($Q < 2.6$, $p >$ adjusted $\alpha$) or interaction between FOD and FOF ($Q <= 4.71$, $p >$ adjusted $\alpha$) on any coherent metrics.

For coherent edits, viewing, read, and feedback, the effect sizes of FOF on the frequency metric were greater than on the proportion metric. Such differences reached 0.11 for coherent edits and 0.14 for coherent feedback. Moreover, there was a

[2] We did not use the WRS2 package because it did not offer the three-way repeated ANOVA.

**Table 2** Descriptive statistics and Kendall's Tau correlations among raw variables

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. FOD | - | | | | | | | | | | | | | | |
| 2. FOF | −0.30* | - | | | | | | | | | | | | | |
| 3. Off-task behavior | 0.09 | 0.03 | - | | | | | | | | | | | | |
| 4. F. Co. edits [a] | −0.19 | 0.06 | −0.36* | - | | | | | | | | | | | |
| 5.F. Co. marking | −0.21* | 0.01 | −0.06 | 0.30* | - | | | | | | | | | | |
| 6. F. Co. viewing | −0.26* | 0.10 | −0.20 | 0.51* | 0.46* | - | | | | | | | | | |
| 7. F. Co. read | −0.15 | −0.03 | −0.25* | 0.33* | 0.16 | 0.32* | - | | | | | | | | |
| 8. F. Co. feedback | −0.24* | 0.07 | −0.27* | 0.50* | 0.31* | 0.46* | 0.39* | - | | | | | | | |
| 9. P. Co. edits [b] | −0.10 | 0.03 | −0.12 | 0.33* | 0.12 | 0.23* | 0.38* | 0.31* | - | | | | | | |
| 10. P. Co. marking | −0.08 | −0.08 | 0.04 | 0.19 | 0.35* | 0.38* | 0.13 | 0.18 | 0.06 | - | | | | | |
| 11. P. Co. viewing | −0.22* | 0.21* | −0.19 | 0.30* | 0.10 | 0.32* | 0.28* | 0.27* | 0.08 | 0.01 | - | | | | |
| 12. P. Co. read | −0.19 | 0.03 | −0.15 | 0.30* | 0.26* | 0.38* | 0.33* | 0.42* | 0.10 | 0.32* | 0.31* | - | | | |
| 13. P. Co. feedback | −0.16 | 0.03 | −0.12 | 0.19 | 0.20 | 0.24* | 0.32* | 0.57* | 0.28* | 0.11 | 0.13 | 0.24* | - | | |
| 14. Pre-test | 0.03 | −0.10 | −0.20 | 0.19 | 0.20 | 0.14 | 0.03 | 0.19 | −0.02 | 0.08 | 0.08 | 0.10 | 0.07 | - | |
| 15. Post-test | −0.14 | 0.05 | −0.22* | 0.40* | 0.36* | 0.28* | 0.13 | 0.33* | 0.16 | 0.24 | 0.11 | 0.21 | 0.20 | 0.34* | - |
| Mean | 3.72 | 2.09 | 0.08 | 0.38 | 0.05 | 0.41 | 0.23 | 0.07 | 0.80 | 0.73 | 0.60 | 0.77 | 0.37 | 4.84 | 8.34 |
| SD | 1.11 | 1.16 | 0.13 | 0.25 | 0.05 | 0.30 | 0.13 | 0.06 | 0.17 | 0.34 | 0.19 | 0.18 | 0.21 | 2.03 | 3.78 |

* $p < \alpha$ adjusted by the Benjamini–Hochberg correction

[a] F. Co. edits, the frequency of coherent edits per minute

[b] P. Co. edits, the proportion of coherent edits

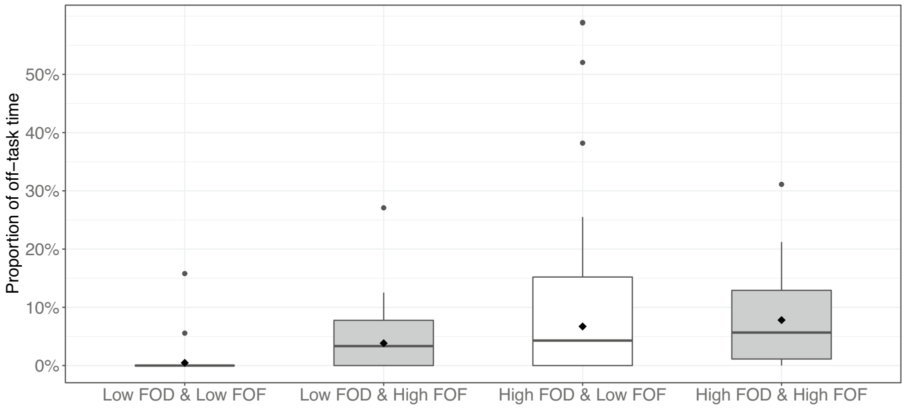


**Fig. 2** The proportion of off-task time in different groups. Rhombuses represent the trimmed means. FOD, feeling of difficulty. FOF, feeling of familiarity

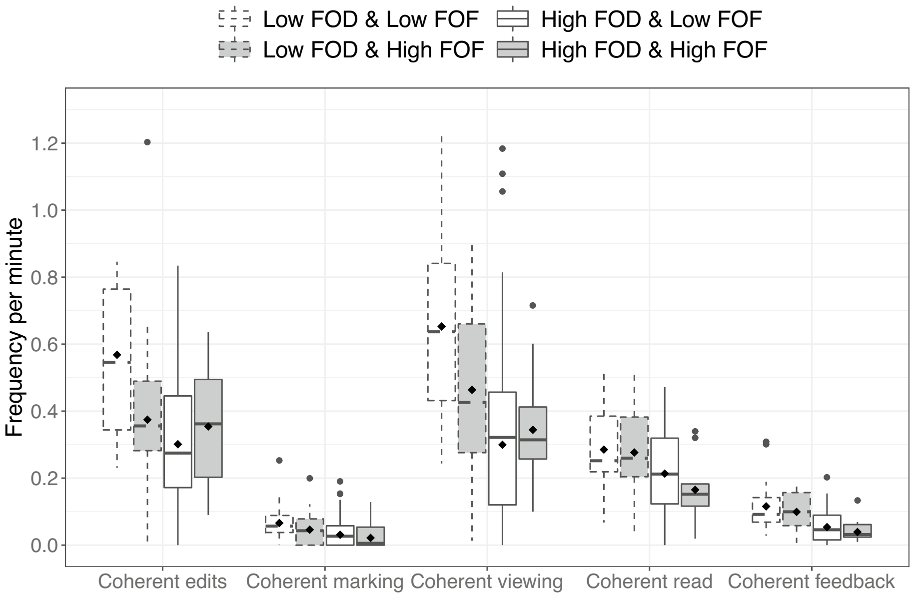


**Fig. 3** The frequency of coherent actions per minute in different groups. Rhombuses represent the trimmed means. FOD, feeling of difficulty. FOF, feeling of familiarity

statistically significant difference in the frequency of coherent edits between high-FOD and low-FOD students but no difference in the proportion of coherent edits. The $\omega_p^2$ of FOF and the interaction was negligible or small on all coherent metrics and was not statistically significant.

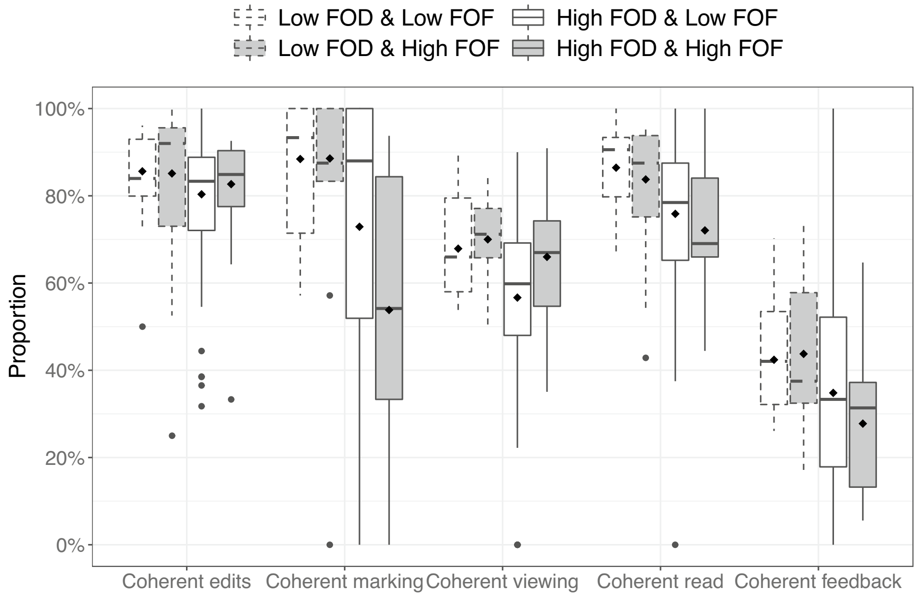


**Fig. 4** The proportion of coherent actions in different groups. Rhombuses represent the trimmed means. FOD, feeling of difficulty. FOF, feeling of familiarity

**Table 3** The robust ANOVA results on the frequency and proportion of coherent actions

| Measures | | FOD | | FOF | | Interaction | |
|---|---|---|---|---|---|---|---|
| | | $Q$ | $\omega_p^2$ | $Q$ | $\omega_p^2$ | $Q$ | $\omega_p^2$ |
| Coherent edits | Frequency | 6.39 | 0.12* | 1.54 | 0.00 | 4.71 | 0.04 |
| | Proportion | 1.11 | 0.00 | 0.06 | 0.00 | 0.15 | 0.00 |
| Coherent marking | Frequency | 5.80 | 0.04* | 1.42 | 0.00 | 0.18 | 0.00 |
| | Proportion | 7.33 | 0.04* | 1.04 | 0.00 | 1.07 | 0.00 |
| Coherent viewing | Frequency | 11.63 | 0.12* | 1.09 | 0.00 | 2.85 | 0.02 |
| | Proportion | 4.60 | 0.09* | 2.60 | 0.01 | 1.03 | 0.00 |
| Coherent read | Frequency | 7.64 | 0.07* | 0.74 | 0.00 | 0.36 | 0.00 |
| | Proportion | 7.72 | 0.05* | 0.65 | 0.00 | 0.02 | 0.00 |
| Coherent feedback | Frequency | 12.93 | 0.17* | 0.84 | 0.01 | 0.00 | 0.00 |
| | Proportion | 6.39 | 0.03* | 0.37 | 0.00 | 0.81 | 0.00 |

*$p <$ adjusted $\alpha$

## Learning

Table 4 displays the means and standard deviations of test scores and their ranks. In all groups, students scored higher in the post-test than in the pre-test. The non-parametric repeated ANOVA revealed a main effect of test time ($F_{ATS,1,\infty} = 108.71$, $p < 0.001$) and a three-way interaction between FOD, FOF, and test time ($F_{ATS,1,\infty} = 5.58$, $p = 0.02$). The

**Table 4** The original test scores and their ranks

| Mean (SD) | Test scores | | Ranks | |
|---|---|---|---|---|
| | Pre | Post | Pre | Post |
| High-FOD high-FOF | 4.20 (1.55) | 9.10 (4.94) | 49.45 (28.07) | 113.6 (39.39) |
| High-FOD low-FOF | 4.73 (1.91) | 7.41 (3.22) | 60.80 (38.63) | 100.23 (43.39) |
| Low-FOD high-FOF | 5.17 (2.62) | 8.90 (4.36) | 66.43 (46.93) | 110.70 (60.30) |
| Low-FOD low-FOF | 4.93 (2.03) | 9.93 (3.58) | 60.79 (37.98) | 129.36 (33.97) |

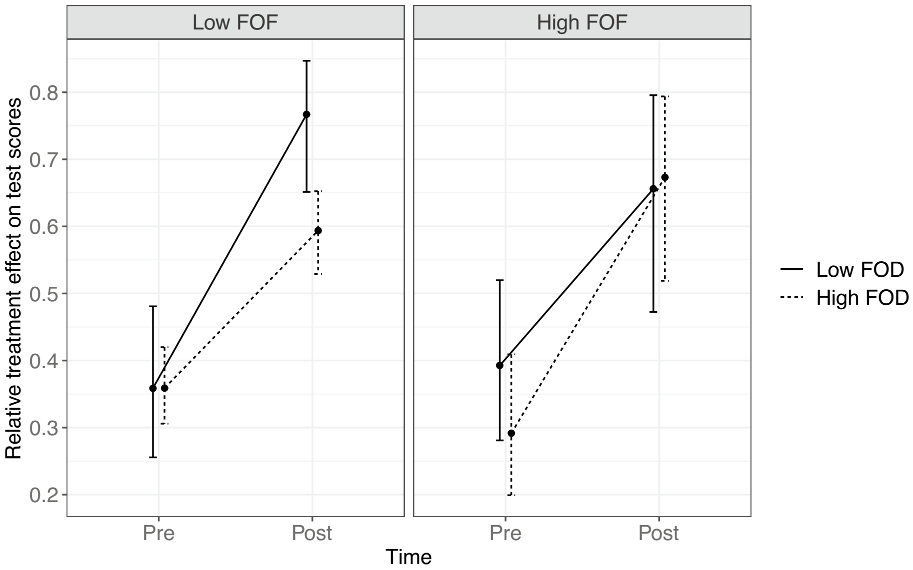


**Fig. 5** The relative treatment effect on test scores with 95% confidence intervals. FOD, feeling of difficulty. FOF, feeling of familiarity

main effects of FOD and FOF, as well as all two-way interactions, were not statistically significant ($F_{ATS,1,\infty} < 1.40$, $p > 0.24$).

Figure 5 depicts the change in relative treatment effects from the pre-test to the post-test, which represented learning gains. Within the high-FOF group, there was no difference in both pre-test and post-test scores between high-FOD and low-FOD students. Within the low-FOF group, high-FOD and low-FOD students also had no difference in the pre-test scores, but a statistically significant difference appeared in the post-test scores. To further investigate the three-way interaction, we conducted a two-way non-parametric repeated ANOVA (FOD × time) within each FOF group. The results showed that there was an interaction between time and FOD within the low-FOF group ($F_{ATS,1,\infty} = 6.27$, $p = 0.01$) but not in the high-FOF group ($F_{ATS,1,\infty} = 1.03$, $p = 0.31$). Thus, within the high-FOF group, learning gains were the same across FOD ratings, but within the low-FOF group, low-FOD students learned more than those with high-FOD.

## Discussion

This study found that FOD was related to various aspects of the learning process, including off-task behaviors, coherent actions, and learning gains. FOF was not related to off-task behaviors and coherent actions, but it moderated the association between FOD and learning gains. In the low-FOF group, students with high-FOD learned less than the low-FOD students, but this pattern did not repeat within the high-FOF group.

The high-FOD group engaged in more off-task behaviors and fewer coherent actions. This result is in line with prior research, where participants read fewer content pages in a difficult information searching task than in an easy task (Liu et al., 2012). FOD is associated with disfluency and may be related to more analytic but slower task processing (Touroutoglou & Efklides, 2010). Thus, high-FOD students might need more time for analyzing collected information and executing coherent actions than low-FOD students. Moreover, FOD indicates cognitive load (Ayres, 2006); when perceived cognitive load is overwhelming, learners may exert little or no effort (Feldon et al., 2019). Due to exerting little effort, high-FOD students might engage in more off-task behaviors than low-FOD students.

High-FOD students also had a lower proportion of coherent actions than low-FOD students. The reason for this relationship may be that FOD is associated with cognitive skills and self-concept (Efklides, 2006), which may impact the application of strategy to regulate learning (Efklides, 2011). However, the difference in the proportion of coherent action was smaller than in the frequency of coherent actions. The inconsistent differences between the proportion and frequency of coherent actions may be explained by the notion that FOD may trigger analytic cognitive processing (Alter et al., 2007; Efklides, 2009). Although the analytic processing is slow, it may alert learners to the discrepancy between progress and goals and trigger the application of metacognitive strategies to regulate learning (Efklides, 2011). Therefore, for high-FOD students, experiencing FOD might mitigate the disadvantage in applying metacognitive strategies caused by factors such as low cognitive skills and self-concept. Nevertheless, the results of this study can only serve as indirect and weak evidence for the claim that FOD may trigger learners' self-regulation (Efklides, 2009), which requires further examination.

In our results, FOF was not related to the proportion of off-task time and coherent action measures. The finding is inconsistent with prior research (Qiu & Lo, 2016; Soemer et al., 2019). One possible reason is related to the specificity of the FOF measurement. In the current research, the learning material was about human thermoregulation and contained 1,974 words and 13 scientific concepts. The FOF item asked students' general familiarity with human thermoregulation rather than the 13 concepts. By contrast, in Soemer et al.'s (2019) study, the learning material was also text about science, but it was much shorter (426 words) and focused on a more limited biological system, the human lung. In Qiu and Lo's study (2016), English learners completed oral narrative tasks within minutes. Examples of task topics were finding a lost item and a job. FOF in Soemer et al.'s and Qiu and Lo's studies was specifically toward the task rather than a more general concept; thus, the measurement of FOF may be more specific in these studies than in the current research. Learners in the current research might recall experiences related to human thermoregulation but not the 13 concepts in mind when they reported FOF. When learners with high FOF actually started the task, they found that they did not know the learning content. In this case, the familiarity-stereotyping effect would not appear because the familiarity-stereotype did not fit the context (Garcia-Marques et al., 2016). This possibility is consistent with the dissociation between FOF and pre-test scores in the current study (see Table 2).

Consequently, high-FOF students might put effort into studying and adopt more analytic processing like their low-FOF counterparts. It is worth noting that the lack of association between FOF and pre-test scores does not mean that the measure of FOF was nonreliable. FOF is the product of a non-analytic inferential process and based on experiences rather than knowledge (Whittlesea, 1993).

The inconsistent findings raise the question about how the overall FOF toward a task is related to FOF toward subtasks, as well as how they are differentially related to learning behaviors and gains. For instance, in Betty's Brain, further research may examine the association between FOF toward single concepts and the overall FOF as well as the difference between the average FOF toward single concepts and the overall FOF. Answers to these questions may inform us about how FOF toward subtasks influence the formation of the overall FOF and how to choose the FOF measurement. Moreover, the difference in the associations between various FOF and learning is also critical because of practical implications. If only FOF specific toward the material can facilitate learning, teachers may need to demonstrate the link between learning material and students' experiences as concretely as possible.

We expected that low-FOD students would learn better than high-FOD students, but this was only true in the low-FOF group. This result may be explained by the mere-exposure effect, which refers to the phenomenon that people tend to rate repeated stimuli likable because of familiarity (Hansen & Wänke, 2009). FOD was negatively associated with joy (Tornare et al., 2015) and situational interest, and positive affect decreased more among students with high-FOD than those with low-FOD over the task process (Fulmer & Tulis, 2013). Within the high-FOF group, the mere-exposure effect might mitigate the undesirable association between affect and FOD by maintaining student interest and positive affect, and thus, lead to no learning difference between low-FOD and high-FOD students.

## Implications

This study's results suggest that FOD may have extensive impacts on learning, including off-task and strategic behaviors and learning gains. However, these results do not imply that teachers should avoid difficult material. The scientific topic of human thermoregulation in this study was challenging for sixth graders: 67% of students rated it as difficult or very difficult. Medium difficulty material, by contrast, may motivate students (Lupo et al., 2019). Moreover, instead of avoiding difficult material, it is crucial to provide cognitive and metacognitive scaffolding to students because such support may mitigate students' FOD toward challenging material (Efklides, 2006). For example, guiding students to set realistic goals may decrease FOD (Guthrie et al., 2013). In Betty's Brain, this may be achieved by supporting students in decomposing the map building task into small subtasks.

This study's findings support the notion that learning material should be linked to students' prior experiences (Rivet & Krajcik, 2008). FOF was not related to any learning behavior metric, but it moderated the relationship between FOD and learning gains. In cases where students lack relevant experiences, teachers may use educational technologies, such as computer-based simulations (D'Angelo et al., 2014; Winn et al., 2006), to familiarize students with the topic before studying the exact content. Nevertheless, some unfamiliarity can be helpful. For instance, unfamiliarity may suppress heuristic and non-analytic information processing via eliminating the familiarity-stereotype effect (Garcia-Marques & Mackie, 2001) and trigger advanced cognitive activities, such as questioning and reasoning

(Sockalingam & Schmidt, 2013). Thus, students' prior experiences and FOF need to be carefully considered while designing learning tasks.

Overall, this study highlights the role of metacognitive experiences in the learning process. Metacognitive knowledge and strategies have been receiving considerable attention, but metacognitive experiences are relatively underexplored. Although this study only investigated the association between metacognitive experiences and behaviors, these experiences also interact with affect and motivation (Efklides, 2009). Examining the interactions among metacognitive experiences, behaviors, affect, and motivation will generate a comprehensive understanding of metacognitive experiences in learning.

## Limitation and further research

FOF and FOD were retrospective in this study. Thus, no causal effect between them and learning behaviors can be concluded from the data available. Further studies may, for instance, present tasks with different difficulty levels to the same individuals and measure both prospective and retrospective FOD toward each task. At the student-level, testing the associations between prospective FOD and learning behaviors examines whether FOD influences learning behaviors. At the task-level, whether individuals apply metacognitive strategies differently when FOD differs provides understanding about the within-students variability in the application of metacognitive strategies across different FOD levels (also see Malmberg et al., 2016).

We operationalized metacognitive strategy use by analyzing the coherence of student actions within Betty's Brain. This definition is conceptually reasonable, since prior studies have found meaningful links between coherent actions, learning, and affect (Zhang et al., 2020; Segedy et al., 2015), but it is worth further investigating the reliability and validity of such definition by triangulating our findings with other methods. For example, future research could use well-established frameworks to collect and code think-aloud data in terms of SRL strategies (Azevedo & Cromley, 2004). In particular, we could use this method to investigate what SRL strategies students' report (via think-aloud) when they execute coherent actions. Understanding the links between coherent actions and SRL strategies may allow finer-grained analyses and enrich the theory of micro-level SRL process (Molenaar, 2014). Moreover, interviews targeted on coherent actions may be conducted in situ to reveal what students are thinking when they execute coherent actions. Thus far, coherence analyses have been limited to Betty's Brain and CTSiM, a computer-based environment for learning scientific phenomena, computational concepts, and practices (Zhang et al., 2021). The extent to which this approach applies to other contexts, such as intelligent tutoring systems and educational games, awaits further investigation.

It may be valuable to measure metacognitive experiences at different time points over the course of learning and examine their dynamic association with behaviors. Results from such analyses may enrich micro-level self-regulated learning theories (Molenaar, 2014). Researchers may measure metacognitive experiences via the daily diary approach or momentary time sampling (Gunthert & Wenze, 2012; Meany Daboul et al., 2007), depending on the length of the task and the granularity of interest. Since having students report metacognitive experiences frequently may intervene learning, a less invasive approach will be helpful. For instance, FOD is associated with some verbal and nonverbal indicators, such as utterance denoting inability to solve the task and frowning (Efklides, 2016). Researchers may collect these data and apply machine learning to build automatic detectors of metacognitive experiences (see, for instance, Ching-En, 2018).

In the current study, FOD and FOF were each measured by one item with five options. This one-item measurement has been used in most prior studies about metacognitive experiences (e.g., Dindar et al., 2020; Efklides, 2002), but it lowers reliability and limits the range of FOD and FOF to five values. Such a small range might cause failure in identifying the optimal FOD, which is analogous to the optimal task difficulty (Malmberg et al., 2013), and optimal FOF. At the optimal level, the benefits of FOD and FOF may be maximized, while their undesirable effects may be minimized. Further research may explore the optimal metacognitive experiences by capturing these constructs in a fine-grained fashion.

## Conclusion

This study investigated the role of FOD and FOF in learning a complex scientific phenomenon, human thermoregulation, within a computer-based environment. It found that students with high FOD demonstrated more off-task behaviors and less strategic behaviors than those with low FOD. However, the difference between the two groups was larger in terms of the frequency of strategic behaviors than its proportion. FOF was related to neither off-task behaviors nor strategic behaviors, but it moderated the association between FOD and learning. In the low-FOF group, students with low FOD learned more than those with high FOD, but this effect did not exist in the high-FOF group.

**Acknowledgements** We thank Anabil Munshi for his help in preprocessing the data and Stefan Slater for his helpful comments on an earlier version of this paper.

**Author contribution** Yingbin Zhang—conceptualization, methodology, formal analysis, writing (original draft preparation, review, and editing). Luc Paquette—conceptualization, methodology, writing (review and editing), supervision. Ryan S. Baker—methodology, writing (review and editing). Nigel Bosch—methodology, writing (review and editing). Jaclyn Ocumpaugh—methodology, investigation, writing (review and editing). Gautam Biswas—methodology, resources, writing (review and editing).

**Funding** This work was supported by the National Science Foundation (grant numbers 1561676) and the China Scholarship Council (grant numbers 201806040180).

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

Alexander, P. A., Kulikowich, J. M., & Jetton, T. L. (1994). The role of subject-matter knowledge and interest in the processing of linear and nonlinear texts. *Review of Educational Research, 64*(2), 201–252. https://doi.org/10.3102/00346543064002201

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*(5), 389–400. https://doi.org/10.1016/j.learninstruc.2006.09.001

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist, 50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Azevedo R., Aleven V. (2013) Metacognition and learning technologies: An overview of current interdisciplinary research. In Azevedo R. & Aleven V. (Eds), *International handbook of metacognition* *and learning technologies* (pp. 1–16). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-5546-3_1

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523–535. https://doi.org/10.1037/0022-0663.96.3.523

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427–449). Springer New York. https://doi.org/10.1007/978-1-4419-5546-3_28

Beckmann, J. F., & Goode, N. (2014). The benefit of being naïve and knowing it: The unfavourable impact of perceived context familiarity on learning in complex problem solving tasks. *Instructional Science, 42*(2), 271–290. https://doi.org/10.1007/s11251-013-9280-7

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's brain. *International Journal of Artificial Intelligence in Education, 26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Borgers, N., Sikkel, D., & Hox, J. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity, 38*(1), 17–33. https://doi.org/10.1023/B:QUQU.0000013236.29205.a6

Brunner, E., & Puri, M. L. (2001). Non-parametric methods in factorial designs. *Statistical Papers, 42*(1), 1–52. https://doi.org/10.1007/s003620000039

Ching-En, C. (2018). Metacognitive experience modeling using eye-tracking. In Penstein Rosé C. et al. (Eds), *Artificial Intelligence in Education. AIED 2018. Lecture Notes in Computer Science, Vol 10948* (pp. 503–507). Cham, Germany: Springer. https://doi.org/10.1007/978-3-319-93846-2_94

Craig, K., Hale, D., Grainger, C., & Stewart, M. E. (2020). Evaluating metacognitive self-reports: Systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning, 15*, 155–213. https://doi.org/10.1007/s11409-020-09222-y

D'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., ... Haertel, G. (2014). *Simulations for STEM learning: Systematic review and meta-analysis*. Menlo Park, CA: SRI International.

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review, 28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed mode surveys: The tailored design method (4th ed.)*. John Wiley & Sons Inc.

Dindar, M., Järvelä, S., & Järvenoja, H. (2020). Interplay of metacognitive experiences and performance in collaborative problem solving. *Computers and Education*, *154*, Article 103922. https://doi.org/10.1016/j.compedu.2020.103922

Efklides, A. (2002). Feelings and judgments as subjective evaluations of cognitive processing: How reliable are they? *Psychology: The Journal of the Hellenic Psychological Society*, *9*(2), 163–184.

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review, 1*(1), 3–14. https://doi.org/10.1016/j.edurev.2005.11.001

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist, 13*(4), 277–287. https://doi.org/10.1027/1016-9040.13.4.277

Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema, 21*(1), 76–82.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645

Efklides, A. (2016). Metamemory and affect. In J. Dunlosky & S. U. K. Tauber (Eds.), *The Oxford handbook of metamemory.* (pp. 245–267). New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.1

Efklides, A., Kourkoulou, A., Mitsiou, F., & Ziliaskopoulou, D. (2006). Metacognitive knowledge of effort, personality factors, and mood state: Their relationships with effort-related metacognitive experiences. *Metacognition and Learning, 1*(1), 33–49. https://doi.org/10.1007/s11409-006-6581-0

Efklides, A., Papadaki, M., Papantoniou, G., & Kiosseoglou, G. (1998). Individual differences in feelings of difficulty: The case of school mathematics. *European Journal of Psychology of Education, 13*(2), 207–226. https://doi.org/10.1007/BF03173090

Efklides, A., Samara, A., & Petropoulou, M. (1999). Feeling of difficulty: An aspect of monitoring that influences control. *European Journal of Psychology of Education, 14*(4), 461–476. https://doi.org/10.1007/BF03172973

Efklides, A., & Tsiora, A. (2002). Metacognitive experiences, self-concept, and self-regulation. *Psychologia, 45*(4), 222–236. https://doi.org/10.2117/psysoc.2002.222

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition. *Perspectives on Psychological Science, 8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review, 31*(2), 319–337. https://doi.org/10.1007/s10648-019-09464-6

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fulmer, S. M., & Tulis, M. (2013). Changes in interest and affect during a difficult reading task: Relationships with perceived difficulty and reading fluency. *Learning and Instruction, 27*, 11–20. https://doi.org/10.1016/j.learninstruc.2013.02.001

Garcia-Marques, T., Mackie, D. M., Maitner, A. T., & Claypool, H. M. (2016). Moderation of the familiarity-stereotyping effect: The role of stereotype fit. *Social Cognition, 34*(2), 81–96. https://doi.org/10.1521/soco.2016.34.2.81

Garcia-Marques, T., & Mackie, D. M. (2001). The feeling of familiarity as a regulator of persuasive processing. *Social Cognition, 19*(1), 9–34. https://doi.org/10.1521/soco.19.1.9.18959

KE Godwin H Seltman M Almeda M Davis Skerbetz S Kai RS Baker AV Fisher 2021 The elusive relationship between time on-task and learning: Not simply an issue of measurement Educational Psychology 1–18 https://doi.org/10.1080/01443410.2021.1894324

Gunthert, K. C., & Wenze, S. J. (2012). Daily diary methods. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 144–159). The Guilford Press.

Guthrie, J. T., Klauda, S. L., & Ho, A. N. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly, 48*(1), 9–26. https://doi.org/10.1002/rrq.035

Hansen, J., & Wänke, M. (2009). Liking what's familiar: The importance of unconscious familiarity in the mere-exposure effect. *Social Cognition, 27*(2), 161–182. https://doi.org/10.1521/soco.2009.27.2.161

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers and Education, 90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

Kinoshita, S. (1997). Masked target priming effects on feeling-of-knowing and feeling-of-familiarity judgments. *Acta Psychologica, 97*(2), 183–199. https://doi.org/10.1016/S0001-6918(97)00018-8

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Liu, C., Liu, J., Cole, M., Belkin, N. J., & Zhang, X. (2012). Task difficulty and domain knowledge effects on information search behaviors. *Proceedings of the American Society for Information Science and Technology, 49*(1), 1–10. https://doi.org/10.1002/meet.14504901142

Lupo, S. M., Strong, J. Z., & Conradi Smith, K. (2019). Struggle is not a bad word: Misconceptions and recommendations about readers struggling with difficult texts. *Journal of Adolescent and Adult Literacy, 62*(5), 551–560. https://doi.org/10.1002/jaal.926

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods, 52*(2), 464–488. https://doi.org/10.3758/s13428-019-01246-w

Malmberg, L., Lim, W. H., Tolvanen, A., & Nurmi, J. (2016). Within-students variability in learning experiences, and teachers' perceptions of students' task-focus. *Frontline Learning Research*, *4*(5), 62–82. https://doi.org/10.14786/flr.v4i5.227

Malmberg, L., Walls, T. A., Martin, A. J., Little, T. D., & Lim, W. H. (2013). Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models. *Learning and Individual Differences, 28*, 54–65. https://doi.org/10.1016/j.lindif.2013.09.007

Meany Daboul, M. G., Roscoe, E. M., Bourret, J. C., & Ahearn, W. H. (2007). A comparison of momentary time sampling and partial-interval recording for evaluating functional relations. *Journal of Applied Behavior Analysis, 40*(3), 501–514. https://doi.org/10.1901/jaba.2007.40-501

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43–59. https://doi.org/10.1007/BF02505024

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*(3), 159–163. https://doi.org/10.1111/j.1467-8721.2009.01628.x

Molenaar, I. (2014). Advances in temporal analysis in learning and instruction. *Frontline Learning Research, 2*(4), 15–24. https://doi.org/10.14786/flr.v2i4.118

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation, 26*, 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the non-parametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, *50*(12), 1–23. https://doi.org/10.18637/jss.v050.i12

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: A meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition and Learning*, *13*(2), 179–212. https://doi.org/10.1007/s11409-018-9183-8

Patall, E. A., Hooper, S., Vasquez, A. C., Pituch, K. A., & Steingut, R. R. (2018). Science class is too hard: Perceived difficulty, disengagement, and the role of teacher autonomy support from a daily diary perspective. *Learning and Instruction, 58*, 220–231. https://doi.org/10.1016/j.learninstruc.2018.07.004

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801–813. https://doi.org/10.1177/0013164493053003024

Qiu, X., & Lo, Y. Y. (2016). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research, 21*(6), 681–698. https://doi.org/10.1177/1362168816684368

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a double-edged sword: A computational model of the encoding/retrieval trade-off with familiarity. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Vol. 48. Skill and strategy in memory use* (pp. 271–312). Elsevier Academic Press. https://doi.org/10.1016/S0079-7421(07)48007-0

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 435–451. https://doi.org/10.1037/0278-7393.18.3.435

Rellinger, E., Borkowski, J. G., Turner, L. A., & Hale, C. A. (1995). Perceived task difficulty and intelligence: Determinants of strategy use and recall. *Intelligence, 20*(2), 125–143. https://doi.org/10.1016/0160-2896(95)90029-2

Rivet, A. E., & Krajcik, J. S. (2008). Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. *Journal of Research in Science Teaching, 45*(1), 79–100. https://doi.org/10.1002/tea.20203

Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2019). Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacognition and Learning, 14*(1), 1–19. https://doi.org/10.1007/s11409-019-09188-6

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science, 43*(1), 93–114. https://doi.org/10.1007/s11251-014-9328-3

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1074–1083. https://doi.org/10.1037/0278-7393.18.5.1074

Segedy, J. R. (2014). *Adaptive Scaffolds in Open-Ended Computer-Based Learning Environments* (Order No. 3674132). [Doctoral dissertation, Vanderbilt University]. ProQuest Dissertations & Theses Global.

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics*, 2(1), 13–48. https://doi.org/10.18608/jla.2015.21.3

Sockalingam, N., & Schmidt, H. G. (2013). Does the extent of problem familiarity influence students' learning in problem-based learning? *Instructional Science, 41*(5), 921–932. https://doi.org/10.1007/s11251-012-9260-3

Soemer, A., Idsardi, H. M., Minnaert, A., & Schiefele, U. (2019). Mind wandering and reading comprehension in secondary school children. *Learning and Individual Differences*, *75*, Article 101778. https://doi.org/10.1016/j.lindif.2019.101778

Soppe, M., Schmidt, H. G., & Bruysten, R. J. M. P. (2005). Influence of problem familiarity on learning in a problem-based course. *Instructional Science, 33*(3), 271–281. https://doi.org/10.1007/s11251-004-7688-9

Tansey, E. A., & Johnson, C. D. (2015). Recent advances in thermoregulation. *Advances in Physiology Education, 39*(3), 139–148. https://doi.org/10.1152/advan.00126.2014

Tornare, E., Czajkowski, N. O., & Pons, F. (2015). Children's emotions in math problem solving situations: Contributions of self-concept, metacognitive experiences, and performance. *Learning and Instruction, 39*, 88–96. https://doi.org/10.1016/j.learninstruc.2015.05.011

Touroutoglou, A., & Efklides, A. (2010). Cognitive interruption as an object of metacognitive monitoring: Feeling of difficulty and surprise. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 171–208). Boston: Springer. https://doi.org/10.1007/978-1-4419-6546-2_9
Tulis, M., & Fulmer, S. M. (2013). Students' motivational and emotional experiences and their relationship to persistence during academic challenge in mathematics and reading. *Learning and Individual Differences, 27*, 35–46. https://doi.org/10.1016/j.lindif.2013.06.003
van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*(1), 16–26. https://doi.org/10.1080/00461520701756248
Veenman, M. V. J., Elshout, J. J., & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction, 7*(2), 187–209. https://doi.org/10.1016/S0959-4752(96)00025-4
Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(6), 1235–1253. https://doi.org/10.1037/0278-7393.19.6.1235
Whittlesea, B. W., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica, 98*(2–3), 141–165. https://doi.org/10.1016/S0001-6918(97)00040-1
Wilcox, R. R. (2011). *Introduction to robust estimation and hypothesis testing (*3rd ed*)*. San Diego, CA: Academic press. https://doi.org/10.1016/B978-0-12-804733-0.00014-7
Winn, W., Stahr, F., Sarason, C., Fruland, R., & Oppenheimer, P.,... Lee, Y. L. (2006). Learning oceanography from a computer simulation compared with direct experience at sea. *Journal of Research in Science Teaching, 43*(1), 25–42. https://doi.org/10.1002/tea.20097
Yang, Z., Barnard-Brak, L., & Lan, W. Y. (2019). Examining the association of over-claiming with mathematics achievement. *Learning and Individual Differences, 70*, 30–38. https://doi.org/10.1016/j.lindif.2019.01.004
Zhang, N., Biswas, G., & Hutchins, N. (2021). Measuring and analyzing students' strategic learning behaviors in open-ended learning environments. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-021-00275-x
Zhang, Y., Paquette, L., Baker, R. S., Ocumpaugh, J., Bosch, N., Munshi, A., & Biswas, G. (2020). The relationship between confusion and metacognitive strategies in Betty's Brain. In C. Rensing & H. Drachsler (Eds.), *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 276–284). Association for Computing Machinery. https://doi.org/10.1145/3375462.3375518

**Luc Paquette.** Department of Curriculum & Instruction, University of Illinois at Urbana-Champaign, 1310 S Sixth Street, Champaign, IL 61820, USA. ORCID: 0000-0002-2738-3190.
**Yingbin Zhang.** Department of Curriculum & Instruction, University of Illinois at Urbana-Champaign, 1310 S Sixth Street, Champaign, IL 61820, USA. E-mail: yingbin2@illinois.edu. ORCID: 0000-0002-2664-3093.

*Current themes of research:*

Our current themes of research are students' self-regulated learning processes during science learning within open-ended environments. Specifically, we focus on the interactions among behaviors, affect, and motivation.

*Most relevant publications in the field of Psychology of Education:*

Paquette, L., Grant, T., Zhang, Y., Biswas, G., & Baker, R. (2021). Using epistemic networks to analyze self-regulated learning in an open-ended problem-solving environment. In Ruis A.R., Lee S.B. (eds), *Advances in Quantitative Ethnography. ICQE 2021. Communications in Computer and Information Science, vol 1312* (pp. 185–201).Cham: Springer. https://link.springer.com/chapter/10.1007/978-3-030-67788-6_13#citeas.
Zhang, Y., Paquette, L., Baker, R. S., Ocumpaugh, J., Bosch, N., Munshi, A., & Biswas, G. (2020).The relationship between confusion and metacognitive strategies in Betty's Brain. In Rensing, C. & Drachsler, H. (Eds), *Proceedings of the 10th International Conference on Learning Analytics & Knowledge* (LAK'20) (pp. 276–284). New York, NY: ACM. https://doi.org/10.1145/3375462.3375518.