

The relationship between confusion and metacognitive strategies in Betty's Brain

Yingbin Zhang
Curriculum and
Instruction

University of Illinois at
Urbana-Champaign
Champaign, IL, USA
yingbin2@illinois.edu

Luc Paquette
Curriculum and
Instruction

University of Illinois at
Urbana-Champaign
Champaign, IL, USA
lpaq@illinois.edu

Ryan S. Baker
Graduate School of
Education

University of
Pennsylvania
Philadelphia, PA, USA
rybaker@upenn.edu

Jaclyn Ocumpaugh
Penn Center for
Learning Analytics

University of
Pennsylvania
Philadelphia, PA, USA
aandres@gse.upenn.edu

Nigel Bosch

School of Information
Sciences

University of Illinois at
Urbana-Champaign
Champaign, IL, USA
pnb@illinois.edu

Anabil Munshi

Institute for Software
Integrated Systems

Vanderbilt University
Nashville, TN, USA
anabil.munshi@vanderb
ilt.edu

Gautam Biswas

Institute for Software
Integrated Systems

Vanderbilt University
Nashville, TN, USA
gautam.biswas@vander
bilt.edu

ABSTRACT

Confusion has been shown to be prevalent during complex learning and has mixed effects on learning. Whether confusion facilitates or hampers learning may depend on whether it is resolved or not. Confusion resolution, behind which is the resolution of cognitive disequilibrium, requires learners to possess some skills, but it is unclear what these skills are. One possibility may be metacognitive strategies (MS), strategies for regulating cognition. This study examined the relationship between confusion and actions related to MS in Betty's Brain, a computer-based learning environment. The results revealed that MS behavior differed during and outside confusion. However, confusion resolution was not related to MS behavior, and MS did not moderate the effect of confusion on learning.

CCS CONCEPTS

• Applied computing~Computer-assisted instruction • Applied computing~E-learning

KEYWORDS

Confusion, confusion resolution, metacognitive strategy, learning analytics

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

LAK '20, March 23–27, 2020, Frankfurt, Germany

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7712-6/20/03...\$15.00

<https://doi.org/10.1145/3375462.3375518>

ACM Reference format:

Yingbin Zhang, Luc Paquette, Ryan Baker, Jaclyn Ocumpaugh, Nigel Bosch, Anabil Munshi, and Gautam Biswas. 2020. The relationship between confusion and metacognitive strategies in Betty's Brain. In *Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK '20)*. ACM, New York, NY, USA, X pages. <https://doi.org/10.1145/1234567890>

1 Introduction

Previous research has suggested that confusion may promote robust learning (e.g., [1-3]). Whether confusion facilitates or hampers learning depends on whether it is resolved or not [3]. If a computer-based learning environment (CBLE) can intentionally confuse learners and provide scaffolding to promote confusion resolution, it may successfully promote deep learning [4]. The scaffolding should have the potential to induce learners' use of strategies to resolve confusion. However, it is unclear what strategies are conducive to confusion resolution.

Metacognitive strategies (MS) – strategies for monitoring and controlling cognitive processes [5] – are an essential component of self-regulated learning (e.g., [6]) and a powerful predictor of academic performance (e.g., [7]). Recent research has raised the possibility that MS might be necessary for confusion resolution [8, 9], and researchers also suggested promoting self-regulation strategies to help learners take advantage of confusion in CBLEs [10]. However, to date, there is little direct evidence to show that self-regulation or metacognitive strategies can improve the chances of confusion resolution.

The research presented in this paper aims to address the gap in the relationship between confusion and MS by examining whether MS are associated with confusion and confusion resolution in the context of Betty's Brain, an open-ended CBLE. The results suggest that the use of MS may not be sufficient for confusion

resolution. This finding has important implications for developing CBLEs that take advantage of confusion. In order to help learners to resolve confusion, CBLEs may need to consider both learners' cognitive and motivational characteristics in addition to facilitating learners' MS behavior.

1.1 Confusion during Learning

Confusion is an emotion related to knowledge [11, 12] and has been widely observed across various learning environments (e.g., [13-15]). According to Mandler's interruption (discrepancy) theory [16], confusion follows cognitive disequilibrium, which arises when there is a conflict between new information and an individual's prior knowledge, or the individual's current knowledge structure cannot assimilate the new information. D'Mello and Graesser [17] hypothesized that confusion is the affective signature of cognitive disequilibrium based on a set of systematic studies [2, 3, 18]. These studies found that confusion occurred more often in conditions that could cause cognitive disequilibrium.

Pekrun and Stephens [12] indicated that cognitive disequilibrium does not initially or inevitably trigger confusion. Learners initially experience surprise and curiosity when new information is incongruent with their knowledge. If this incongruity cannot be resolved, then confusion arises. If the incongruity seems to be impossible to dissolve, frustration replaces confusion.

Silvia [11] supposed that events stemming from appraisals of high novelty and low comprehensibility lead to confusion, while events with appraisals of high novelty but high comprehensibility induce interest. Following Silvia's claim, Muis, Chevrier, and Singh [19] argued that confusion more likely follows surprise when the task is quite complicated. This notion was supported by [20]. They found that the relative frequency of confusion was higher than curiosity when learners felt that information was novel and complex.

To summarize, discrepant information or novel and complex information causes cognitive disequilibrium, which leads to confusion if it cannot be resolved right away.

Effects of confusion on learning. The effects of confusion on learning are mixed. On the negative side, confusion has been associated with less frequent use of deep learning strategies and planning strategies [8, 9]. Uninterrupted confusion might become frustration [21], and this may cause students to feel lower levels of self-efficacy [22]. This, in turn, can lead to lower achievement [8] or even more serious disengagement, including dropping out of courses [13, 23].

On the positive side, confusion can promote students' MS behavior [9, 24] and benefit learning [2, 3]. D'Mello and Graesser [17] claimed that confusion "plays a prominent role in learning activities that are pitched at deeper levels of comprehension" (pp. 290). The extent to which learners can benefit from confusion may depend on both their skills and their motivation. For instance, one study found that students with higher cognitive ability and

drive had higher learning gains resulting from confusion than their peers [25].

Confusion Resolution. The mixed effects of confusion on learning may be related to its resolution. Poor learning outcomes may be partially due to persistent unresolved confusion [17, 26]. However, confusion need not to be entirely resolved in order to be beneficial for learning; both completely and partially resolved confusion have been observed to predict learning gains [3].

These results could help explain the conflicting impact of confusion on learning. When confusion is entirely or partially resolved, learning outcomes improve, but when confusion is entirely unresolved, learners are unable to modify their existing knowledge structure and assimilate the new information. Instead, they learn nothing, become frustrated [21] and doubt their abilities [22]. Therefore, it is critical to understand how confusion can be resolved so that intervention and scaffolding can be designed and given to students who are likely to experience sustained confusion.

While this paper discusses confusion resolution, it is important to clarify that confusion resolution does not mean learners directly resolving their confusion. Rather, what they tackle is the cognitive disequilibrium, the cause of confusion [2, 3]. In other words, confusion resolution excludes situations where learners simply ignore their confusion and engage in the next task.

D'Mello et al. [2] indicate that confusion can be resolved if (1) learners have the knowledge and skills to resolve it, or (2) the learning environment provides scaffolding to help them resolve it. The assumption of the first prerequisite was supported by another study [3], where those with higher ACT (a standardized college entrance test) scores were more likely to partially resolve their confusion in learning the functioning of everyday devices, such as the electric bell and the toaster. These prerequisites suggested that confusion resolution requires special skills, and a try at resolving confusion without the necessary skills is insufficient to lead to its resolution [27].

1.2 Metacognitive Strategies

Metacognitive strategies (MS), or metacognitive skills, are procedural knowledge of how, when, and why to regulate cognition [5, 28]. Generally, its components include goal setting, planning, self-monitoring, self-control, and self-evaluation [29].

Impact of MS on Learning. Many studies have found the benefits of MS for learning. In an early literature review, MS had been revealed to be one of the most powerful predictors of academic performance [30]. Some subsequent studies did not find a strong association between MS and learning (e.g., [24, 31]), but a recent meta-analysis, which focused on studies in elementary and secondary school, indicated that the correlation between learning and MS differed significantly across specific components of MS, disciplines, grades and how learning and MS are measured [29]. For example, the average Pearson product-moment correlations of the two variables were 0.21, 0.23, 0.26, and 0.34 in math, English/language arts, science, and social studies, respectively. The results were replicated in another meta-analysis

[7], which also found that the relationship between MS and learning was moderated by disciplines, grades, and the measurement approach of MS.

1.3 Confusion and Metacognitive Strategy

Nelson and Narens' [32] two-level metacognitive system provides a helpful framework for understanding the relationship between cognitive disequilibrium and MS. It proposes the object-level, where cognition about the external world is, and the meta-level, where cognition about cognition is. The meta-level receives information from the object-level to monitor the cognitive activities and send instructions to the object-level to control cognitive activities. The controlling function of the meta-level manifests as the use of MS [28]. Under the two-level metacognitive system, learners need to activate MS to resolve confusion. When learners encounter discrepant information, this information flows from object-level to the meta-level, and the meta-level detects the inconsonant state in the object-level. To restore cognitive equilibrium, the meta-level must modify cognitive activities in the object-level through the controlling function. In other words, learners should behave according to MS so that they can integrate the discrepant information and their existing knowledge model.

The metacognitive and affective model of self-regulated learning (MASRL; [6]) also informs the association between confusion and the use of MS. The MASRL model emphasizes the interactions of affect, motivation, and metacognition in self-regulated learning (SRL). Under this model, cognitive

disequilibrium during task processing leads to affective reactions, such as surprise and confusion. At the same time, cognitive disequilibrium also contributes to metacognitive experiences (ME), such as the feeling of difficulty and the estimate of effort and time. The ME and affective states, in turn, may trigger decisions about conducting MS behavior. In other words, cognitive disequilibrium may indirectly influence the use of MS. Therefore, it is reasonable to expect that learners change MS behavior when they are confused.

Taking these theories together, the changes in MS behavior may co-occur with confusion because both follow cognitive disequilibrium. In order to make cognition return to the state of equilibrium, learners need to control cognitive activities via conducting MS. However, this notion needs further research to examine.

1.4 Current Study

The current study combines the observations of emotion and the action logs of students' activities in a CBLE, Betty's Brain, to answer three research questions examining the relationship between confusion and MS. As described above, confusion may be accompanied with changes in MS behavior. Besides, confusion is experienced as an unpleasant emotion [33] and may motivate learners to regulate their cognition to resolve it. This motivates the first research question (RQ1): Is the frequency of MS behaviors different when students are confused compared with when they are not confused?

Betty's Brain - The Teachable Agents Group @ Vanderbilt University

The screenshot displays the 'Quiz Results' tab for a 'Human Activity Quiz' taken on Saturday, September 28 at 10:19 PM. The quiz score is 17%. Below the results, a concept map titled 'The Concept Map used for this Quiz' illustrates the relationships between various factors. The map shows 'vehicle use' leading to 'fossil fuel use' (increases), which leads to 'carbon dioxide' (increases). 'deforestation' leads to 'vegetation' (decreases), which also leads to 'carbon dioxide' (increases). The 'carbon dioxide' node is highlighted in yellow.

#	Question	Answer	Grade
1.	If vehicle use increases, then what happens to carbon...	carbon dioxide will increase .	✓
2.	If deforestation increases, then what happens to carb...	carbon dioxide will decrease .	✗
3.	If electricity generation increases, then what happens ...	I don't know	?
4.	If factories increase, then what happens to carbon dio...	I don't know	?
5.	If garbage and landfills increase, then what happens t...	I don't know	?

Quiz Score: 17%
The Concept Map used for this Quiz

```

graph LR
    A[vehicle use] -- increases (+) --> B[fossil fuel use]
    B -- increases (+) --> C[carbon dioxide]
    D[deforestation] -- decreases (-) --> E[vegetation]
    E -- increases (+) --> C
    style C fill:#ffff00
  
```

Figure 1. Screenshot of viewing quiz results and checking the chain of links Betty used to answer a quiz question.

The use of MS may be necessary for learners to restore cognition to an equilibrium state and resolve confusion. To study this hypothesis, the current study investigated the second question (RQ2): Is confusion resolution related to increases in MS behaviors?

Confusion can benefit learning when it is resolved or partially resolved [3, 17]. If MS contribute to confusion resolution, the effect of confusion on learning may depend on students' MS. For students experience the same level of confusion, those with high MS may learn better than those with low MS. Thus, the third research question (RQ3) asked: Can MS moderate the relationship between confusion and learning?

2 Betty's Brain

Betty's Brain is an open-ended CBLE [34]. Students learn about scientific phenomena, such as climate change, thermoregulation or river ecosystems, by teaching a virtual pedagogical agent named Betty. They achieve this by building a causal map of the scientific phenomenon, in which causal (*cause-and-effect*) relationships are represented by a set of concepts connected by directed causal links (see Figure 1). To build this map, students have access to hypermedia resource pages (*Science Book* in Figure 1) on relevant scientific concepts. Students can evaluate their causal modeling progress by asking Betty to take graded quizzes or by querying her on cause-and-effect questions related to what she has been taught so far. Betty's quiz grades or her explanations to her answers can help the student keep track of her, and in turn their own progress. By looking at Betty's correct and incorrect answers, students can identify problems in their causal map. They can then improve their understanding of the topic by correcting those problems on their causal map.

While using Betty's Brain, students' activities can be grouped into six primary categories, including *read* (read a page in the resources), *note* (create, view or edit a note), *edit* (edit the causal map, including adding or deleting a concept, adding, deleting or modifying a causal link, marking a causal link as correct or incorrect), *query* (ask Betty a cause-and-effect question based on the concepts and links on her map so far), *quiz* (assess the state of the map by having Betty take a quiz), and *explain* (ask Betty to explain her answer to a cause-and-effect query or check the chain of links she used to answer a quiz question, in order to probe Betty's reasoning).

3 Methods

3.1 Participants and Procedures

The data in this study were collected from 93 sixth-grade students in an urban public school in Tennessee, during the 2016-2017 school year. The study lasted seven school days. On day 1, students spent 30-45 minutes on completing a paper-based pre-test that assessed their knowledge of climate change and causal relationships. On day 2, they received a 30-minute training about

how to use Betty's Brain. In the next four days, they spent 45 to 50 minutes per day teaching Betty about climate change by constructing the causal map. On the final day, students completed a post-test identical to the pre-test.

3.2 Observations of Emotion

While students were working on Betty's Brain, two trained observers recorded their affective states via the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP; [35]). The protocol was enforced by the Android App HART [36], which time stamps each observation. With the timestamps, the observations could be aligned with the log file of student actions in Betty's Brain.

The observers used a momentary time sampling method [37], where they coded students individually in a pre-determined order. This sampling method aims at obtaining a representative sample of emotions and behavior. The observers recorded both students' emotions (i.e., boredom, confusion, delight, engaged concentration, and frustration) and behavior (i.e., on-task, on-task conversation, off-task), though the current study only considers the emotion observations. They recorded the first thing they saw and had up to 20 seconds to decide. The inter-rater reliability was good (Cohen's Kappa > 0.60). The observations of confusion were 326, and there were 233, 157, 4182, and 251 observations of boredom, delight, engaged concentration, and frustration, respectively. On average, each student had 53 observations.

3.3 Metacognitive Strategy

We analyzed the action logs of students working on Betty's Brain to measure MS behavior based on coherence analysis (CA) [38]. CA interprets learners' behavior according to the relationship between actions (e.g., read and edit). Two ordered actions, *x* and *y*, are interpreted as being coherent if the second action, *y*, utilizes information generated by the first action, *x*, because *x* supports *y*. For instance, in Figure 1, the quiz results could tell students the question about the causal relation of deforestation and carbon dioxide was answered incorrectly. After viewing these quiz results, if students read the resource pages that contained information about the correct relationship between the two concepts, the action of viewing quiz results supported the action of reading resource pages, and they were coherent. Although *x* and *y* need not be two consecutive actions, the time interval between them should be restricted to avoid coincidental connections. In line with prior work in Betty's Brain (e.g., [38]), we set the restriction for the current study to 5 minutes.

CA assumes that action coherence may act as evidences for metacognitive skills and task understanding [38]. Indeed, if a student shows two ordered actions that are coherent, it suggests they monitored previous cognitive processes (*the first action*, e.g., viewing quiz results), received the information generated by previous cognitive processes (*the results of the first action*, e.g., the link chain between deforestation and carbon dioxide was wrong), and regulated the later cognitive processes (*the second action*, e.g., reading resource pages about deforestation and carbon dioxide) based on the received information. Therefore,

performing coherent actions may suggest that a student engages in MS behavior.

Five behavioral metrics were identified as measures of MS. Each metric was calculated as the frequency of a particular kind of action per minute so that these metrics were comparable across students, given that the total time of each student's working on Betty's Brain varied slightly. Except for the first metric, the other four were based on CA.

1. *Quizzing* frequency: the number of times the student had Betty take a quiz, per minute. This variable reflects how often students evaluated Betty's understanding, i.e., their own understanding. This metric is not based on CA, but taking tests is often thought as the self-monitoring strategy for evaluating current progress (e.g., [39]). For simplicity, we include this metric in our list of CA metrics.
2. Frequency of *coherent viewing*: the number of viewing quiz results actions, per minute, that were coherent with later actions. This variable measures how often student collected assessment information that was used later. Coherent viewing might indicate self-monitoring behavior because information it generated supported subsequent actions.
3. Frequency of *coherent editing*: the number of coherent map edits per minute. It measures how often students edited the concept map based on acquired information. Coherent editing might indicate self-control behavior because the editing action depended on the results of monitoring processes, i.e., the information generated by reading resource pages or viewing quiz results.
4. Frequency of *coherent reading*: the number of coherent reading page actions per minute. This variable measures how often students intentionally sought relevant information to improve their understanding based on the quiz results. Coherent reading might represent self-control behavior because the page being read was contingent on the quiz results.
5. Frequency of *coherent marking*: the number of coherent actions of marking (i.e. labeling) a link on the concept map, per minute. In addition to marking these links as "correct" or "maybe wrong", students can also delete these marks. This variable reflects how often, based on the quiz results, students understood what links on their map were correct or possibly incorrect, and annotated them accordingly. Coherent marking might represent self-control behavior because the marking action depended on the quiz results. It might also indicate the use of self-monitoring strategy because the mark could remind students of which parts of their understanding were incorrect, correct, or uncertain.

In line with prior research on Betty's Brain (e.g., [38]), we excluded some actions that were too short, including *coherent viewing* (quiz results) that were less than 2 seconds and *coherent reading* (resource pages) actions that were less than 10 seconds. These actions were discarded because they are likely to indicate only shallow engagement with that part of the system. For example, short reading actions may indicate that students are just browsing through the resource pages quickly without reading the text.

3.4 Knowledge tests

The pre-test and post-test were identical in both questions and forms. The test assessed students' knowledge of climate change and causal relationships. It contained multiple-choice questions and short-answer questions. The answer to each short-answer question consisted of a fixed number of successive steps. Students could get maximum scores of 7 points and 9 points for the multiple-choice questions and the short-answer questions, respectively, and a total maximum score of 16 points.

3.5 Analyses

RQ1. CA metrics derived from the action logs generated within the duration of an emotion were regarded as measures of MS behavior during this emotion. Although the field observations were conducted using a 20-second time window, we analyzed the data with a window including 30 seconds before and 30 seconds after the observation of the emotion, based on prior research that found the average duration of emotions during learning is approximately 30 seconds [21, 40] and because it is impossible to know whether an observation was collected closer to the beginning or the end of a learner's emotion. For this study, CA metrics were calculated for 326 observations of confusion and 4,823 observations of other emotions.

The Shapiro-Wilk test indicated that the distribution of the CA metrics was skewed ($p < 0.001$ for all metrics). Therefore, the Wilcoxon test was used to compare CA metrics during confusion and outside confusion (during other emotions). The Benjamini-Hochberg correction (FDR) was applied to adjust the p -value and control for false discovery rate.

RQ2. If confusion is resolved, or the cognitive disequilibrium is resolved, the learner's affective state is expected to transit to engaged concentration [21, 40] or delight. Otherwise, confusion continues, and persistent confusion can lead to frustration or boredom [21, 40]. Liu and colleagues [41] also found that prolonged confusion was negatively related to learning, while short-term confusion was positively related to learning. Therefore, if an observation of confusion is followed by an observation of engaged concentration or delight ("confusion \rightarrow engaged concentration/delight"), this confusion may be resolved. In contrast, if an observation of confusion is followed by an observation of confusion, frustration, or boredom ("confusion \rightarrow confusion/frustration/boredom"), this confusion may be unresolved.

The current study refers to a pair of successive emotion observations for the same student as an affect sequence, which we categorize based on valance. Specifically, positive affect sequences referred to "confusion \rightarrow engaged concentration/delight", while negative affect sequences referred to "confusion \rightarrow confusion/frustration/boredom". The affect sequences whose intervals were longer than 3 minutes were discarded because students might experience another affective state between two successive observations of affective states if the interval between the two observations was too long (e.g., [21, 40]). Three minutes was used as the cut-off value because it corresponded to the average gap between one observation and the

next one.¹ In order to answer RQ2, the Wilcoxon test was conducted to compare CA metrics during confusion in positive affect sequences and negative affect sequences. The Benjamini-Hochberg correction was applied.

RQ3. The five CA metrics were calculated for each student across their whole action logs. Principal component analysis (PCA) with varimax rotation was applied to extract a factor from the five CA metrics. This factor was used as an overall indicator of MS. Multicollinearity was tested to examine whether the five metrics were suitable for PCA. Table 1 displays the correlation matrix. Seven of the ten Pearson product-moment correlations were greater than 0.4. Bartlett's test of sphericity indicated that sufficient correlations existed among the CA metrics ($\chi^2 / df = 18.80, p < 0.001$).

Table 1. Correlations among coherence analysis metrics.

	1	2	3	4
1. Quizzing	-			
2. Coherent viewing	0.53	-		
3. Coherent editing	0.39	0.57	-	
4. Coherent reading	0.66	0.69	0.6	-
5. Coherent marking	0.20	0.55	0.25	0.41

The Kaiser-Meyer-Olkin (KMO) overall statistic was 0.76, indicating that a high proportion of variance in the CA metrics might be caused by underlying factors, and all single KMO values were greater than 0.5 (Table 2). Cronbach's alpha for these metrics was 0.83, indicating that the internal consistency of the CA metrics was good. Overall, the results suggested that these metrics were suitable for PCA. Table 2 displays the results of PCA.

Table 2. The results of the principal component analysis.

<i>M</i> (<i>SD</i>)	MSA	Loading	R ²
Quizzing	0.75	0.73	54%
Coherent viewing	0.78	0.88	77%
Coherent editing	0.80	0.74	55%
Coherent reading	0.76	0.89	79%
Coherent marking	0.69	0.59	35%
Total explained variance			60%

¹ The gap varies because the time that observers needed to decide an observation was not constant. For example, if one student showed apparent confusion (e.g., scrunching up the nose and forehead and pursing lips), observers might only need 5 seconds to conclude that this student's emotion was confusion; in another observation where this student might just purse lips slightly, observers might need 15 seconds to make a decision.

Students' confusion scores were calculated as the ratio of the number of confusion observations to the total number of affect observations. Students were divided into high and low confusion groups, based on whether their confusion scores were higher or lower than the median, as well as high and low MS groups, based on whether their MS scores were higher or lower than the median. Data from ten students were discarded because of missing either their pre-test or post-test scores. Finally, there were 23 in the high-MS and high-confusion group, 20 in the high-MS and low-confusion group, 18 in the low-MS and high-confusion group, and 22 in the low-MS and low-confusion group.

The Shapiro-Wilk test indicated that the distribution of test scores was skewed ($p = 0.007$), and thus, a three-way non-parametric mixed ANOVA was conducted to answer research question 3. The test time was the within-subject factor, and confusion and MS were the between-subject factors. The analysis was implemented within the *nparLD* package in R [42].

4 Results

4.1 Is the Frequency of MS Behavior Different during and outside the Duration of Confusion?

Table 3 shows the results of the Wilcoxon test. The frequencies of *coherent reading* and *viewing* were higher during confusion than when students were in other affective states. This result indicates that the information that students collected while they were confused was more likely to be used to support later actions. Similarly, they were more likely to read pages based on quiz results. There were no statistically significant differences for quiz-taking actions, coherent map editing actions, and coherent link marking actions. These results suggest that learners changed their use of MS when confused.

Table 3. Coherence analysis metrics during and outside confusion.

<i>M</i> (<i>SD</i>)	During	Outside	Cohen's <i>d</i>
Quizzing	0.13 (0.37)	0.16 (0.41)	0.08
Coherent viewing	0.52 (1.11)	0.26 (0.71)	0.19***
Coherent editing	0.26 (0.56)	0.32 (0.67)	0.10
Coherent reading	0.38 (0.66)	0.26 (0.54)	0.20**
Coherent marking	0.05 (0.28)	0.04 (0.36)	0.01

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.2 Is Confusion Resolution Related to Increases in MS Behaviors?

Table 4 displays the results of the Wilcoxon test. There was no difference in the frequency of CA metrics during confusion in negative and positive affect sequences (adjusted $p > 0.2$ for all metrics).

Table 4. Coherence analysis metrics in negative and positive affect sequences.

<i>M</i> (<i>SD</i>)	Negative (<i>N</i> = 26)	Positive (<i>N</i> = 63)
Quizzing	0.23 (0.43)	0.08 (0.27)
Coherent editing	0.35 (0.56)	0.25 (0.59)
Coherent reading	0.19 (0.40)	0.32 (0.69)
Coherent marking	0.04 (0.19)	0.05 (0.37)
Coherent viewing	0.65 (1.32)	0.59 (1.50)

4.3 Can MS Moderate the Relationship between Confusion and Learning?

Table 5 displays the pre-test and post-test scores of different groups. For participants with low MS, the difference between pre-test and post-test scores was greater in the low-confusion group (Cohen's $d = 1.03$) than in the high-confusion group (Cohen's $d = 0.88$). In contrast, among the high-MS group, this difference was smaller in the low-confusion group (Cohen's $d = 0.99$) than in the high-confusion group (Cohen's $d = 1.22$). However, the non-parametric mixed ANOVA only revealed a marginally significant interaction among time, confusion, and MS ($F = 3.11, p = 0.078$). The main effects of time and MS were significant ($F = 98.18$ and $F = 16.47$, respectively; both $p < 0.001$). Students performed better in the post-test than in the pre-test, and students with high MS had greater test scores than the low-MS group. The main effect of confusion was not significant ($F = 1.89, p = 0.169$). No significant interaction was found between time and MS ($F = 1.51, p < 0.22$), time and confusion ($F = 2.53, p = 0.11$), as well as MS and confusion ($F = 0.00, p = 0.950$).

Table 5. The pre-test and post-test scores.

MS	Confusion	M (SD)		<i>d</i>
		Pre-test	Post-test	
Low	Low	5.84 (2.56)	8.23 (3.19)	1.03
	High	4.97 (3.00)	7.31 (2.41)	0.88
High	Low	8.18 (2.67)	10.28 (2.5)	0.99
	High	6.46 (3.12)	10.41 (3.42)	1.22

5 Discussion

Confusion can benefit learning when it is partially or entirely resolved (e.g., [3]). To resolve confusion, learners need necessary knowledge and skills or need to receive scaffolding or support from the environment [2]. What skills may contribute to confusion resolution is unclear. This study examined the relationship between MS and confusion in the context of Betty's Brain.

5.1 MS and Confusion

Results for RQ 1 showed that the use of MS differed depending on whether the student was experiencing confusion. The results were in line with the MASRL model [6], which hypothesizes that cognitive disequilibrium indirectly impacts the use of MS via

triggering affective reactions and metacognitive experience. Specifically, this study found that two MS strategies, coherent reading and coherent viewing were more common among students when they were confused.

Both behaviors may serve to help learners to find the discrepant information, which is the cause of cognitive disequilibrium [16]. For example, in this study, the correct link between vehicle use and carbon dioxide is "vehicle use increases fossil fuel use, which in turn increases carbon dioxide". Students might think vehicle use could directly increase carbon dioxide and miss the mediator, fossil fuel use, and thus, they would add a link of "vehicle use increases carbon dioxide". When the quiz result indicated the link was wrong, they might feel confused. This unpleasant emotion might drive them to investigate why the link was wrong, and thus they would look for and read the resource pages demonstrating the relationship between the vehicle use and carbon dioxide.

5.2 MS and Confusion Resolution

RQ 2 investigated the relationship between confusion resolution and MS at a fine grain-size, i.e., whether there is more MS behavior during resolved confusion than during unresolved confusion? Results showed that the frequency of MS behavior was not significantly different during the confusion followed by engaged concentration or delight (i.e., resolved confusion) and during the confusion followed by confusion, frustration, or boredom (i.e., unresolved confusion). This suggests that MS behavior may not be sufficient for resolving the cognitive disequilibrium that underlies confusion. However, it may also be possible that, in some cases, students were cycling to a positive affect like engaged concentration or delight because they were shifting their focus to other scientific concepts and causal relations between these concepts that were less confusing, without resolving the cognitive disequilibrium. Results for RQ 1 suggests that MS may help learners find the discrepant information, for example, the resource page containing information about the wrong link. However, in order to successfully integrate the information into their mental model, students needed to interpret it correctly, and this task might rely on their reading skills.

5.3 MS, Confusion Resolution, and Learning

RQ 3 examined confusion resolution and MS at a coarser granularity. The significant main effect of test time indicated that students' understanding of the thermoregulation increased after using Betty's Brain. This learning was not associated with MS and confusion because the interactions between test time and MS, as well as test time and confusion, were not significant.

However, a marginally significant interaction among test time, confusion, and MS was found. For the low-MS group, students experiencing high confusion had less learning gains than those experiencing low confusion. On the contrary, for high-MS group, students experiencing high confusion had greater learning gains than those experiencing low confusion. The different effects of confusion on learning between high and low MS groups imply that confusion resolution may be associated with MS. Students

with high MS might be likely to resolve confusion, while students with low MS might be unable to resolve confusion. Resolved confusion can benefit learning [2, 3], but unresolved confusion negatively influences learning [17, 26]. Thus, within the low-MS group, confusion might hamper learning, while within the high-MS group, confusion might facilitate learning.

Nevertheless, the interaction among time, confusion and MS was small and only marginally statistically significant. Several reasons might explain this result. Firstly, MS may not be sufficient for resolving the cognitive disequilibrium underlying confusion. As mentioned in 5.2, some cognitive skills may be necessary.

Moreover, in terms of self-regulated learning (SRL), motivation may also be necessary for confusion resolution because it plays a critical role in the regulation of emotion and behavior [6, 43]). This may also partially explain why prior studies did not find a treatment effect for MS scaffolding on learning in Betty's Brain (e.g., [44]). Future work should examine comprehensive motivational, cognitive, and metacognitive data to deeply investigate how these factors influence the resolution process of confusion.

Moreover, confusion did not frequently occur in the current study. Even within the high confusion group, the average proportion of confusion in the affect observations was only 6.0%. The interaction between confusion and time was also not significant, indicating that confusion generally might not benefit learning in this study. Therefore, even though MS might contribute to confusion resolution, their interaction did not exist in this study. It is worth noting that in most prior studies that found the positive effect of confusion, confusion was induced intentionally with discrepant information (e.g., [2, 3, 18]). In contrast, confusion appeared naturally in this study. Therefore, the discrepancy of new information encountered by students might not be as strong as in that earlier work. Most confused states might not reach the zone of optimal confusion, where the positive effect of confusion occurs [2, 10]. Thus, in further research about confusion resolution, it may be more effective to trigger confusion purposely rather than letting it occur naturally. Nevertheless, if a finding can only be obtained through intentionally triggered confusion, it raises questions about the generality and scope of the phenomenon.

Another reason for our relative weak results may be that the proficiency of sixth graders' MS in this study was not strong enough for resolving confusion. The metacognitive skills of children at this age are still developing [45]. Muis et al. [8] also suggested that pupils might not possess necessary skills for confusion resolution.

5.4 Implications

The association between confusion and the change of MS behavior implies that cognitive disequilibrium may not only contribute to metacognitive experience [28] and confusion [16] but also indirectly influence the use of MS. This provided evidence for some assumptions of MASRL model [6] because,

according to this model, the cognitive interruption can trigger metacognitive experience and affective reaction, which in turn motivate bottom-up self-regulation of behavior and emotions. The bottom-up self-regulation contrasts with top-down self-regulation, which is driven by the goal and plan that learners set based on person characteristics such as competency, self-concept, and beliefs concerning learning.

Researchers have suggested taking advantage of confusion in CBLEs by providing self-regulation strategy support [10]. However, this study indicates that MS alone may be not sufficient for confusion resolution. Therefore, if researchers want to promote deep learning via strategy scaffolding for confusion resolution, both cognitive and motivational factors should be considered. If learners lack necessary cognitive skills and intrinsic or extrinsic motivation, strategic scaffolding may fall.

6 Conclusion

This study examined the relationship between confusion and metacognitive strategies. The results showed that changes in MS behavior co-occur with confusion, but that confusion resolution was not related to MS behavior. Furthermore, MS did not moderate the effect of confusion on learning. These results demonstrate that MS may be a prerequisite but not sufficient for confusion resolution.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation under Grant #1561676 and the China Scholarship Council under Grant #201806040180.

REFERENCES

- [1] Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3), 241-250.
- [2] D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153-170.
- [3] D'Mello, S., & Graesser, A. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychologica*, 151, 106-116.
- [4] Lehman, B., & Graesser, A. (2014). Impact of agent role on confusion induction and learning. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS 2014)*, Springer, Honolulu, HI, USA, 45-54.
- [5] 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.
- [6] Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, 46(1), 6-25.
- [7] 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.
- [8] Muis, K. R., Psaradellis, C., Lajoie, S. P., Di Leo, I., & Chevrier, M. (2015). The role of epistemic emotions in mathematics problem solving. *Contemporary Educational Psychology*, 42, 172-185.
- [9] Di Leo, I., Muis, K. R., Singh, C. A., & Psaradellis, C. (2019). Curiosity... Confusion? Frustration! The role and sequencing of emotions during mathematics problem solving. *Contemporary Educational Psychology*, 58, 121-137.

- [10] Arguel, A., Lockyer, L., Kennedy, G., Lodge, J. M., & Pachman, M. (2019). Seeking optimal confusion: a review on epistemic emotion management in interactive digital learning environments. *Interactive Learning Environments*, 27(2), 200-210.
- [11] Silvia, P. J. (2010). Confusion and interest: The role of knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts*, 4(2), 75-80.
- [12] Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. Harris, S. Graham, T. Urdan, S. Graham & J. Royer (Eds.), *Individual Differences and Cultural and Contextual factors. APA Educational Psychology Handbook. Vol. 2* (pp. 3-31). Washington, DC: American Psychological Association.
- [13] Yang, D., Kraut, R., & Rose, C. P. (2016). Exploring the effect of student confusion in massive open online courses. *Journal of Educational Data Mining*, 8(1), 52-83.
- [14] D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*, 105(4), 1082-1094.
- [15] Rodrigo, M., & Baker, R. (2011). Comparing learners' affect while using an intelligent tutor and an educational game. *Research and Practice in Technology Enhanced Learning*, 6(1), 43-66.
- [16] Mandler, G. (1990). Interruption (discrepancy) theory: Review and extensions. In S. Fisher & C. L. Cooper (Eds.), *On the move: The Psychology of Change and Transition* (13-32). Chichester: Wiley.
- [17] D'Mello, S., & Graesser, A. (2014). Confusion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (289-310). New York: Routledge.
- [18] Lehman, B., D'Mello, S., & Graesser, A. (2012). Confusion and complex learning during interactions with computer learning environments. *The Internet and Higher Education*, 15(3), 184-194.
- [19] Muis, K. R., Chevrier, M., & Singh, C. A. (2018). The role of epistemic emotions in personal epistemology and self-regulated learning. *Educational Psychologist*, 53(3), 165-184.
- [20] Chevrier, M., Muis, K. R., Trevors, G. J., Pekrun, R., & Sinatra, G. M. (2019). Exploring the antecedents and consequences of epistemic emotions. *Learning and Instruction*, 63, 101209.
- [21] D Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145-157.
- [22] Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., ... Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology*, 100(3), 525-534.
- [23] Yang, D., Wen, M., Howley, I., Kraut, R., & Rose, C. (2015). Exploring the effect of confusion in discussion forums of massive open online courses. In *Proceedings of the Second ACM Conference on Learning @ Scale (L@S '15)*. ACM, New York, NY, USA, 121-130.
- [24] Muis, K. R., Pekrun, R., Sinatra, G. M., Azevedo, R., Trevors, G., Meier, E., ... Heddy, B. C. (2015). The curious case of climate change: Testing a theoretical model of epistemic beliefs, epistemic emotions, and complex learning. *Learning and Instruction*, 39, 168-183.
- [25] Lehman, B., D Mello, S., & Graesser, A. (2013). Who benefits from confusion induction during learning? An individual differences cluster analysis. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED 2013)*. Springer, Memphis, TN, USA, 51-60.
- [26] Bosch, N., D Mello, S., & Mills, C. (2013). What emotions do novices experience during their first computer programming learning session? In *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED 2013)*. Springer, Memphis, TN, USA, 11-20.
- [27] Lehman, B., & Graesser, A. (2015). To resolve or not to resolve? That is the big question about confusion. In *Proceedings of the 17th International Conference on Artificial Intelligence in Education (AIED 2015)*. Springer, Madrid, Spain, 216-225.
- [28] Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3-14.
- [29] 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.
- [30] Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research*, 84(1), 30-43.
- [31] Hargrove, R. A., & Nietfeld, J. L. (2015). The impact of metacognitive instruction on creative problem solving. *The Journal of Experimental Education*, 83(3), 291-318.
- [32] Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, 26, 125-173.
- [33] Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223-241.
- [34] 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.
- [35] Ocumpaugh, J., Baker, R. S., & Rodrigo, M. M. T. B. (2015). *Baker Rodrigo Ocumpaugh monitoring protocol (BROMP) 2.0 technical and training manual*. Technical Report. New York, NY: Teachers College, Columbia University. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.
- [36] Ocumpaugh, J., Baker, R. S., Rodrigo, M. M., Salvi, A., Van Velsen, M., Aghababayan, A., ... Martin, T. (2015). *HART: The human affect recording tool*. In *Proceedings of the 33rd Annual International Conference on the Design of Communication (SIGDOC '15)*. ACM, New York, NY, USA, Article 24, 6 pages.
- [37] 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.
- [38] 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.
- [39] Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., ... Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. In *Proceedings of the 16th International Conference on Learning Analytics & Knowledge (LAK'20)*. ACM, New York, NY, USA, 191-200.
- [40] Botelho, A. F., Baker, R. S., Ocumpaugh, J., & Heffernan, N. T. (2018). Studying affect dynamics and chronometry using sensor-free detectors. In *Proceedings of the 11th International Conference on Educational Data Mining (EDM 2018)*. International Educational Data Mining Society, Buffalo, NY, USA, 157-166.
- [41] Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. S. (2013). Sequences of Frustration and Confusion, and Learning. In *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)*. International Educational Data Mining Society, Memphis, TN, USA, 114-120.
- [42] Noguchi, K., Gel, Y. R., Brunner, E., & Konietzschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, 50(12), 1-23.
- [43] Azevedo, R., Behnagh, R., Duffy, M., Harley, J., & Trevors, G. (2012). Metacognition and self-regulated learning in student-centered learning environments. In D. Jonassen & S. Land (Eds.), *Theoretical foundations of student-centered learning environments* (2ed., pp. 171-197). New York: Routledge.
- [44] Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and learning*, 9(2), 187-215.
- [45] van der Stel, M., & Veenman, M. V. (2014). Metacognitive skills and intellectual ability of young adolescents: A longitudinal study from a developmental perspective. *European Journal of Psychology of Education*, 29(1), 117-137.