

# Moments of Confusion in Simulation-Based Learning Environments

Sadia Nawaz<sup>1</sup>, Gregor Kennedy<sup>2</sup>, James Bailey<sup>3</sup>, Chris Mead<sup>4</sup>

## Abstract

Confusion is an important epistemic emotion because it can help students focus their attention and effort when solving complex learning tasks. However, unresolved confusion can be detrimental because it may result in students' disengagement. This is especially concerning in simulation environments using discovery-based learning, which puts more of the onus for learning on the students. Thus, students with misconceptions may become confused. In this study, the possible moments of confusion in a simulation-based predict-observe-explain (POE) environment were investigated. Log-based interaction patterns of undergraduate students from a fully online course were analyzed. It was found that POE environments can offer a level of difficulty that potentially triggers some confusion, and a likely moment of students' confusion was the observe task. It was also found that confidence in prior knowledge is an important factor that can contribute to students' confusion. Students mostly struggled when they discovered a mismatch between the subjective and objective correctness of their responses. The effects of such a mismatch were more pronounced when confusion markers were analyzed than when students' learning outcomes were observed. These findings may guide future works to bridge the knowledge gaps that lead to confusion in POE environments.

## Notes for Practice

- Confusion is an epistemic emotion that is unlikely to be avoided in complex learning tasks.
- Confusion can promote a deeper understanding of the concepts because its resolution often requires critical thinking, inquiry, and effortful problem-solving.
- Simulation-based predict-observe-explain (POE) environments can promote a degree of difficulty that can potentially confuse students.
- One of the likely moments of students' confusion is during the observe phase of the POE learning design.
- Students with misconceptions or incorrect prior knowledge are more likely to become confused during the observe phase than students with correct prior knowledge.
- Students' confidence in their prior knowledge can strongly influence how confused they become.
- Students who have high confidence but make incorrect responses are more likely to show signs of confusion—their effort associated with confusion seems to positively influence their learning outcomes.

## Keywords

Confusion, frustration, confidence, hypercorrection, simulation-based learning, predict-observe-explain, affect, metacognitive mismatch

**Submitted:** 28/01/2020 — **Accepted:** 13/10/2020 — **Published:** 17/12/20

Corresponding author <sup>1</sup>Email: [nawazs@student.unimelb.edu.au](mailto:nawazs@student.unimelb.edu.au) Address: School of Computing and Information Systems (CIS), University of Melbourne, Parkville VIC 3052, Australia, ORCID ID: <https://orcid.org/0000-0002-3674-2108>

<sup>2</sup>Email: [gek@unimelb.edu.au](mailto:gek@unimelb.edu.au) Address: Centre for the Study of Higher Education, University of Melbourne, Parkville VIC 3052, Australia, ORCID ID: <https://orcid.org/0000-0001-9423-7786>

<sup>3</sup>Email: [baileyj@unimelb.edu.au](mailto:baileyj@unimelb.edu.au) Address: School of Computing and Information Systems, University of Melbourne, Parkville VIC 3052, Australia, ORCID ID: <https://orcid.org/0000-0002-3769-3811>

<sup>4</sup>Email: [chris.mead@asu.edu](mailto:chris.mead@asu.edu) Address: School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA, ORCID ID: <https://orcid.org/0000-0003-4547-6830>

## 1. Introduction

Students in higher education are now routinely offered the opportunity to learn in digital environments and technology-enhanced courses. While these environments can offer scalability (such as MOOCs) and flexibility (with the provision of

anytime, anywhere access), they often require students to be more autonomous in terms of exploring and covering the content of courses and keeping track of their own progress. This may be particularly true for simulation-based learning environments, in part because of their discovery-based nature. In order to effectively learn in these environments, students need to have a certain level of motivation and self-regulated learning skills (Huang, 2002). Students need to negotiate any technical challenges these digital learning environments pose, while at the same time, they need to grapple with new, complex information in the absence of instructors' immediate feedback (Mansour & Mupinga, 2007). There is a high chance for students to become confused while handling these multiple demands (Pachman, Arguel, & Lockyer, 2015).

Confusion is an "epistemic emotion" (Pekrun, Goetz, Titz, & Perry, 2002; Pekrun & Stephens, 2012) that, unlike basic emotions (such as "happy," "sad," or "angry"), is particularly prevalent in educational settings. Confusion is typically experienced as an affective response to the cognitive processing of complex material. It has often been referred to as a "cognitive feeling" or a feeling about one's knowledge state (Clore, 1992; Storm & Storm, 1987). Researchers have also defined confusion as a "noticeable lack of understanding" (Baker, D'Mello, Rodrigo, & Graesser, 2010; D'Mello, Lehman, Pekrun, & Graesser, 2014), or as a feeling of uncertainty (D'Mello et al., 2005), where students do not know what to do next or how to proceed (Keltner & Shiota, 2003; Rozin & Cohen, 2003).

If students' confusion is sustained, it can lead to frustration and boredom (Chiu, Hong, & Dweck, 1997; D'Mello & Graesser, 2014b; D'Mello & Graesser, 2012a; Kostyuk, Almeda, & Baker, 2018; Rodrigo et al., 2007). Under these conditions, students' goal-blockages and misconceptions become reinforced rather than getting corrected (Arguel & Lane, 2015; Arguel, Lockyer, Lipp, Lodge, & Kennedy, 2017). Such a persistent or prolonged state of confusion is associated with negative learning experiences that can lead to students' disengagement (Baker et al., 2010; D'Mello & Graesser, 2012b; D'Mello et al., 2014).

While confusion can be detrimental to learning, not all episodes of confusion are alike (D'Mello et al., 2014). In certain situations, it has been found to positively affect students' learning outcomes by promoting a deeper understanding of the concepts (Craig, Graesser, Sullins, & Gholson, 2004; D'Mello & Graesser, 2014a; D'Mello et al., 2014). Some researchers have argued that in order to resolve their confusion, students need to engage in adaptive learning behaviours, such as making inferences, reflecting on and integrating new knowledge, and reviewing their prior conceptions (Graesser, Ozuru, & Sullins, 2010). Moreover, some learning designs have been proposed as beneficial because they present students with difficulties that can induce cognitive conflict or disequilibrium (Graesser & Olde, 2003). This can potentially confuse the learners, making these educational frameworks effective for students' learning (Ballantyne & Bain, 1995). Examples of such learning designs are refutation texts (Tippett, 2010), desirable difficulties (Bjork, 2011), productive failure (Kapur, 2016), and impasse-driven learning (Brown & VanLehn, 1980; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). Another learning design that has the potential to trigger confusion in students is the simulation-based predict-observe-explain paradigm (White & Gunstone, 1992). Predict-observe-explain (POE), as the name suggests, is a three-phase, iterative learning design (Dalziel, 2010).

1. During the *prediction* phase, students formulate a hypothesis and are often asked to provide their reasons for committing to a particular hypothesis. It has been suggested that questions that draw on explanatory reasoning may be diagnostic of deeper comprehension (Graesser, Baggett, & Williams, 1996).
2. During the *observation* phase, students test their hypothesis by changing parameters or variables in a simulation so they can see the effects of their manipulations. The *observation* phase is crucial for providing students with important insights into their prior-held ideas and beliefs (Driver, 1983). This phase is especially critical for students who propose an incorrect hypothesis, because they can then see a mismatch between their predictions and their observations.
3. Finally, in the *explain* phase, clarifications are provided to students detailing the relationship between variables or parameters that represent the conceptual phenomenon under investigation. This phase aims to help students reconcile any discrepancies between what they predicted and what they observed in the simulation (Gunstone & White, 1980).

A POE learning design can be applied in a range of learning contexts—face to face, online, computer labs, and so on—and can be used to promote students' discussion (White & Gunstone, 1992) and improve their understanding of science concepts (Bowen & Haysom, 2014). Computer-mediated and multimedia-based POE tasks have been found to promote peer-learning opportunities, and other POE tasks have the potential to support conceptual change by probing students' prior knowledge and beliefs (Gunstone, 1990; Tao & Gunstone, 1999). Researchers have encouraged the use of POE-based learning designs and have called them an "inevitable way" of making science courses more interesting and engaging (Karamustafaoğlu & Mamlok-Naaman, 2015; Kibirige, Osodo, & Tlala, 2014; Sreerekha, Arun, & Sankar, 2016).

While POE learning designs have been promoted, the ways in which students make inferences, reflect on their new knowledge, and then assimilate that knowledge while reviewing their prior conceptions has not received much attention. These behaviours of reflecting upon and reviewing prior conceptions have also been associated with students' struggle and confusion resolution behaviours (D'Mello & Graesser, 2014a; D'Mello et al., 2014). Therefore, students' confusion can play an important role in developing their understanding within the POE environments.

It is important to mention that confusion has been studied in different learning environments by a large number of researchers. For example, Bosch, D’Mello, and Mills (2013); Lee, Rodrigo, Baker, Sugay, and Coronel (2011); and Rodrigo et al. (2009) have analyzed the emotions experienced by novices in programming contexts. They report that students’ confusion could impact their learning outcomes and their eventual success in a computer science class. Confusion has been analyzed in a mathematical context where students needed to solve algebra problems (Lagud & Rodrigo, 2010). Confusion has also been the focus of research in visual art and aesthetic research (Silvia, 2009, 2010); in tutoring sessions using the AutoTutor system, where students needed to work on difficult questions related to computer literacy (D’Mello & Graesser, 2011); in chemistry, using virtual laboratory software (Baker et al., 2011); in narrative-centred learning environments (McQuiggan, Robison, & Lester, 2010); and in many others. Confusion has, however, not received much attention in POE-based environments. Therefore, this study capitalizes on a POE learning environment that focuses on the educational model of conceptual simulation. Specifically, it aims to identify how the POE educational paradigm can guide analytics work in the prediction of confusion.

Detecting students’ confusion is “both critical and complex” (Graesser, 2011). It is an essential step in identifying optimally effective behaviours within digital learning environments. Understanding confusion using learning analytics will help researchers understand how students complete assigned tasks in self-directed environments, and when real-time feedback and support may be provided to guide them, with the possibility of improving their affect, enhancing their engagement, and advancing their learning (Baker et al., 2012). Detection of confusion can be particularly useful for helping students avoid persistent or prolonged confusion, which has been associated with poor learning outcomes and performance (D’Mello, Person, & Lehman, 2009; Lee et al., 2011).

## 2. Background

Regarding emotions in a learning context, McLeod (1989) suggests that given a difficult problem, students’ reactions may include many emotions. A major source of these emotions is an interruption (Mandler, 1984), a cognitive conflict (Piaget, 1952b), an impasse (Brown & VanLehn, 1980; VanLehn et al., 2003), or a dissonance (Festinger, 1957) in students’ plans or their planned behaviours. When such an interruption or dissonance arises, students cannot complete tasks as they normally would, and they often experience physiological arousal. The arousal is then interpreted as an emotion (Mandler, 1984; McLeod, 1989).

To describe and understand the events that lead to students’ confusion, D’Mello & Graesser (2010) have used the cognitive disequilibrium theory (Piaget, 1952b, 1985). According to this theory, when students are exposed to complex information, contradictions, anomalies, novelties, or misconceptions or when they encounter a discrepant event where their observations of a phenomenon are inconsistent with their expectations, cognitive disequilibrium is triggered. Cognitive disequilibrium is of critical importance in students’ comprehension and learning processes (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005). Graesser and D’Mello (2012) suggest that “there may be a causal relationship between cognitive disequilibrium and deep learning, with confusion playing either a mediating or moderating role in the process.” Therefore, cognitive disequilibrium has been attributed as a “key signature” (D’Mello, Taylor, & Graesser, 2007), a “prime candidate” (Graesser & D’Mello, 2012), and a “vital element” (Lodge, Kennedy, Lockyer, Arguel, & Pachman, 2018) that may lead to students’ confusion. Additionally, cognitive disequilibrium theory suggests that a persistent or prolonged state of confusion could result in students’ being stuck—often attributed as a state of frustration. Frustration has also been called “hopeless” confusion, where students have a more impoverished understanding of the subject matter and lack available resolution plans (D’Mello & Graesser, 2014a; D’Mello & Graesser, 2012a). So while dissonance or disequilibrium within an “optimal” zone may be associated with confusion, its prolonged state is more likely to result in frustration (Arguel, Lockyer, Kennedy, Lodge, & Pachman, 2019; D’Mello & Graesser, 2014b).

Within the context of the POE framework, an opportunity for students’ learning is created as a result of contradictions or inconsistencies between their *predictions* and their *observations* (Liew & Treagust, 1998). These inconsistencies between students’ observations and their expectations can be useful motivation for their learning (Fensham & Kass, 1988) and information gathering (Markant, Settles, & Gureckis, 2016). They could result in the onset of cognitive conflict or disequilibrium (Fensham & Kass, 1988) and, presumably, confusion during the *observation* task. Students who attend to their confusion and engage in effortful problem solving may be able to resolve their learning impasses and inconsistencies as they complete the *observation* task—a likely occurrence of “productive” or “hopeful” confusion (D’Mello & Graesser, 2012a). For the other students, who are unable to resolve their confusion, chances are that their learning impasses from the *observation* task would sustain, turning their confusion into frustration as they arrive at the *explanation* task—a possible onset of “vicious” confusion (D’Mello, Picard, & Graesser, 2007).

Kennedy and Lodge (2016) conducted a study to analyze students’ affective experience within a simulation-based POE environment. Students’ sessions were captured using screen-recording software. At the end of the session, students were asked

to complete a video-stimulated recall. The findings suggest that students mostly reported being confused during the *observation* phase, and none of the students felt frustrated at this task. As these students progressed to the *explanation* tasks, those who had resolved their learning impasses from the *observation* phase appeared to be “engaged,” whereas others “became frustrated and disengaged by the time they reached the explanation screens” (Kennedy & Lodge, 2016). Given the cognitive disequilibrium theory (Piaget, 1985) and the previous research (Kennedy & Lodge, 2016), it would be expected that within POE environments the most likely *moment* of students’ confusion is during the *observation* phase.

The research hypothesis stating that students are most likely to experience confusion within the *observation* phase of a POE environment leads to the question of what factors may influence these moments of confusion. Poor knowledge in the content area is clearly a factor that can lead to students’ confusion (Graesser et al., 2005; Pachman et al., 2015). Previous research has established that poor prior knowledge or misconceptions can act as “central obstacles” that impede learning of new information (Duit & Treagust, 2003; Kendeou & Broek, 2005). Research has also established that it is not only the level of knowledge that can affect students’ learning behaviours but also their confidence in the knowledge they hold (Kulhavy, 1977; Kulhavy, Yekovich, & Dyer, 1976). Therefore, in addition to prior knowledge, confidence may be important in understanding the role of confusion in POE-based learning environments.

An area of research that relates directly to confusion and confidence in learning is *hypercorrection*. It refers to the situation when students having high confidence in their knowledge or response are provided with feedback suggesting that they are incorrect (Butterfield & Metcalfe, 2001, 2006). A number of studies (Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009, 2010; Metcalfe & Finn, 2011) have shown that the *hypercorrection* effect occurs due to enhanced attention arising as a result of a metacognitive mismatch. This mismatch arises when people find an inconsistency between the subjective and objective correctness of their responses. Thus, a correct response made with low confidence or an incorrect response made with high confidence could result in a metacognitive mismatch. Researchers have suggested that the metacognitive mismatch could impact students’ learning processes and outcomes.

A number of scholars have explored how students with high and low confidence respond to making errors and getting feedback (Anderson, Kulhavy, & Andre, 1971; Butler, Karpicke, & Roediger, 2008; Butterfield & Metcalfe, 2001, 2006; Kulhavy, 1977; Kulhavy & Stock, 1989; Kulhavy et al., 1976). For example, it has been suggested that learners tend to spend the longest time on feedback when high-confidence responses turn out to be incorrect (Anderson et al., 1971). That extra time on feedback could help them identify the inconsistencies between their existing knowledge and the new information (Kulhavy, 1977; Kulhavy & Stock, 1989). While for errors with high confidence, students’ response could be surprise or confusion, the response to low-confidence errors is acceptance, leading students to spend less time on feedback (Kulhavy et al., 1976). Further, from a cognitive neuroscience perspective, Butterfield and Metcalfe (2001, 2006) found that high-confidence errors were more likely to be corrected on a subsequent retest than errors made with low confidence.

### 3. Research Questions

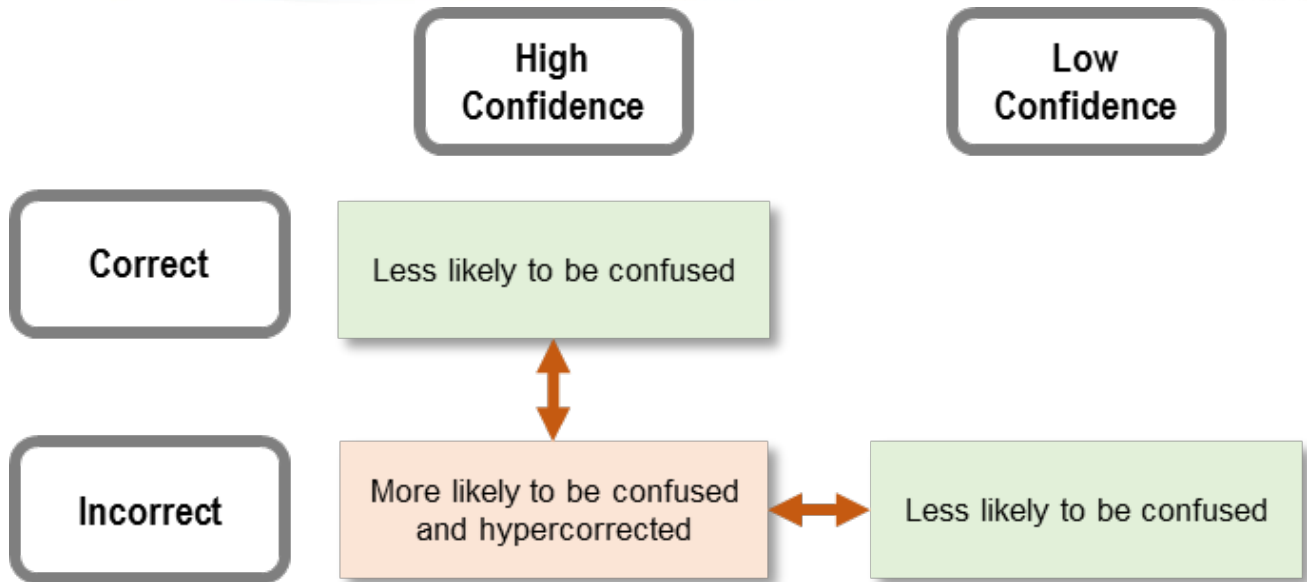
Given these previous research findings, the first general hypothesis investigated in this study was that students who make an incorrect prediction would be more likely to feel confused during the *observation* phase than those who make a correct prediction. Second, and more specifically, it was expected that students who were high in confidence and made an incorrect prediction would show more signs of confusion (during the *observation* phase) than students who were

1. low in confidence and made a similarly incorrect prediction or
2. high in confidence but made a correct prediction.

It should be noted that because of the exploratory nature of this study and the design of the POE learning environments, the hypercorrection effect was investigated in two directions: (1) between the high-confidence correct-predicting students and the high-confidence incorrect-predicting students, and (2) between the high-confidence incorrect-predicting students and the low-confidence incorrect-predicting students (see Figure 1). Because it is less clear from the literature whether students who make correct responses with lower confidence show any signs of confusion, this analysis is not presented in this paper.

Last, it was expected that students who were likely to experience confusion due to a hypercorrection effect (confident but erroneous) would perform better than the other students (less confidence and erroneous, high confidence and correct) on a task that assessed knowledge and understanding.





**Figure 1:** Research hypothesis for the likely moments of students’ confusion in a POE-based simulation environment. It is expected that the students who get confused and *hypercorrected* may develop a better understanding of concepts.

## 4. Methods

### 4.1. Learning Environment

The learning environment in this study was a conceptual simulation contained within a fully online course called Habitable Worlds, which is offered to undergraduate students over eight weeks (Horodyskyj et al., 2018). Habitable Worlds is an introductory science course that encourages students’ problem solving and reasoning through interactive tasks and is built using Smart Sparrow—an “adaptive” e-learning platform that records students’ interactions and activities. *Adaptive* in this context could mean learning-pathway modifications or provision of personalized feedback and hints on students’ responses (or lack of responses).

The Habitable Worlds course consisted of 67 *modules*, which were mostly built around “training” and “application” *tasks*. A task or screen in the current context could be one or more of the following: answering questions through drop-down menus, responding to multiple-choice questions (MCQs), writing free-text answers to the questions, making hypotheses, watching short lecture-style videos, or making submissions associated with simulations. On a given screen as students hit the “submit” or “continue” button to progress to the next task, it was recorded in the system as a task attempt.

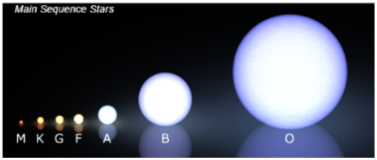
Each “training” module introduced a new concept. These modules were mostly linear in structure, with occasional pathway adaptivity for remediation of learners with prior misconceptions (Pardos & Horodyskyj, 2019). Students could not proceed in a training module unless they correctly completed the current task. Moreover, students could not go back to a previously completed task without restarting the module. “Application” modules comprised quizzes. No new content was introduced, but there was a requirement for students to show competency in the topics that had already been covered.

The primary focus of this paper was on an initial POE-based training module called Stellar Lifecycles. This module was made available to students during the second week of the course, and it was assumed that most students would complete it within that week. Overall, this module consisted of 23 different tasks or screens, but this study only focused on a subset of POE-related tasks. Additionally, the associated application module of Stellar Lifecycles, called Stellar Applications, was included as knowledge-transfer task. This module (transfer task) was available to students for only one week. In Stellar Lifecycles the key concept students were asked to investigate was the relationship between the mass of a star and its associated lifespan. The main sequence of POE-based activities for this module is provided below.

1. During the *prediction* phase, students needed to hypothesize the possible relationship between a star’s mass and its lifespan (see Figure 2). They were also required to state the reasons they believed in their proposed hypothesis. Notably, students were not provided with content relating to this concept before making predictions.

### Stellar Lifecycles

How do you think mass and stellar lifetime are related for Main Sequence stars?

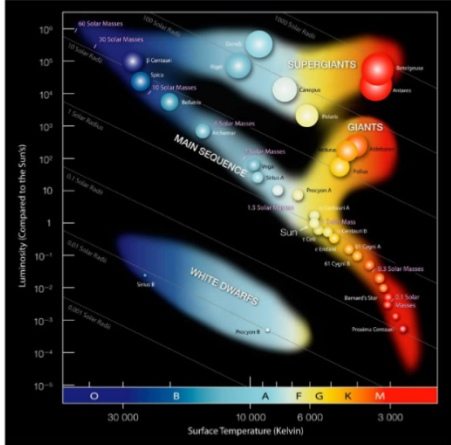


**Main Sequence Stars**

M K G F A B O

- There is no relationship between mass and lifetime of a star.
- Low-mass stars live longer than high-mass stars.
- Medium-mass stars live longer than either low-mass and high-mass stars.
- Medium-mass stars live shorter than either low-mass and high-mass stars.
- High-mass stars live longer than low-mass stars.

Why do you think this is the case?



**TEST HYPOTHESIS**

**Figure 2:** *Predict* task: Students are required to select a hypothesis and state the reasons they believe in that hypothesis.

2. The *observation* phase was divided into two separate tasks.
  - (a) The first task introduced the stellar simulator so that students could learn how to manipulate stars of varying solar mass and how to run a simulation about them. Students could run the stellar simulator as many times as they wanted.
  - (b) The second task asked students to create and run simulations about stars by varying their mass. Then they needed to record the mass and lifespans of these stars in the space provided (see Figure 3). Last, students were asked to either endorse or reject their earlier proposed predictions. Again, they could run the simulation as many times as they wished.

### Stellar Lifecycles

Compare Elements Recycle

CREATE STAR + CLEAR RESET



TIME: 800,280 years SPEED: x 10<sup>5</sup>

Results

Mass: 0.08 Ms	Lifetime: 5783221965271 years
Mass: 12 Ms	Lifetime: 20975547 years
Mass: 25 Ms	Lifetime: 3348239 years

Mass vs. Lifetime Relationship

You proposed: (untested)

**VERIFY**

**Figure 3:** *Observe* task: Students engaged in an interactive task; they needed to create and run virtual stars and then record their observations of stellar mass and associated lifespans. In the end, they needed to either endorse or reject their prior predictions.

3. The *explain* phase was divided into three tasks.
  - (a) The first task was only available to students who made incorrect predictions and endorsed them or those who made correct predictions but rejected them. This task aimed to help students rectify their hypotheses.
  - (b) The second task asked students to report on the minimum and the maximum lifespan of seven different stellar classes. For this aspect of the task, students were asked to create and run simulations about stars of a specified mass range and then to report on the associated lifespans. This task aimed to promote further understanding of the concepts by allowing students to reconcile or resolve their cognitive conflict.
  - (c) The third task explained concepts through a short lecture-style video. This task summarized students' earlier proposed predictions, their newly formed predictions, and their evaluation of the lifespans of the seven stellar classes. Students were then tested on some of the new content from the video.
4. After the *POE* task, students made observations of how, as a star aged, it changed its stellar classification. They were then asked to respond to MCQs about the observed stellar classes.
5. Finally, a *knowledge-transfer* task was presented to students, asking them to calculate the properties of six stars (properties such as luminosity, temperature, and mass) and to identify the longest-lived and the shortest-lived stars. The maximum achievable score was 10 for completely correct answers. While students could make multiple attempts at this task, they were penalized by two marks for each incorrect attempt.

In Figures 2 and 3, some visual information is presented to students through the Hertzsprung–Russell (H–R) diagram. It depicts the zones that a dying star may pass through and does not provide direct information on how a star's mass may be related to its lifespan. Therefore, within the current context, the H–R diagram is “interesting but superfluous” (Kennedy & Lodge, 2016).

#### 4.2. Participants

This study utilized the data from the Spring 2016 offering of the course Habitable Worlds. In accordance with an approved institutional review board (IRB) protocol, student data was anonymized. The participants were 364 non-science major undergraduate students from a large US-based university. Of the students who attempted the Stellar Lifecycles module within the Habitable Worlds course, 51% were female and 49% were male.

Based on age, 44% of students were 20 years old or younger, 38% were older than 20 and younger than 30, and the remaining 18% were older than 30. Based on the academic level, 9% were in their first year, 34% were in their second, 33% were in their third, and the remaining 24% were in their fourth.

#### 4.3. Measures

In this naturalistic study, the majority of analyses focused on the *observe* phase of the POE learning design—given that this is where *moments* of confusion were expected, as predicted by theory (Fensham & Kass, 1988; Graesser & D’Mello, 2012; Liew & Treagust, 1995, 1998; Lodge et al., 2018; Piaget, 1985). Furthermore, the previous study (Kennedy & Lodge, 2016) within the same simulation environment suggested that most students experience confusion during the *observation* phase. This was based on students' affective experiences through video-stimulated recall (Kennedy & Lodge, 2016). This paper aims to extend these previous works to explore whether learning analytics can help identify students' moments of confusion.

This study used trace data to interpret students' behaviours relating to confusion. Trace data or log files are created when students take actions within the digital or online environments, for example, when they watch a video, when they make a submission associated with tasks or questions, when they open a module, or when they close a learning session by signing out of the system (Bunderson, Inouye, & Olsen, 1989; Greiff et al., 2014). Task attempts were used as a first measure of analyzing confusion. This measure was calculated by aggregating the numerical count of task-related submissions—referred to as “count data” (Kovanovic et al., 2015). This data is useful because it can provide an overview of students' learning-based activities. The second parameter used to measure confusion in this study was students' time on task. Research suggests that time on task can be an “accurate” measure for analyzing students' effort on learning (Kovanovic et al., 2015).

Some studies have analyzed students' confusion in terms of their interaction patterns. Baker et al. (2012), in their work on Cognitive Tutor Algebra I, discussed students' features of confusion. They described how the confused students seemed to struggle, becoming slower (i.e., taking more time) on tasks after making two or more errors. These students made errors more frequently and were also found to be less sure and guess more in their responses. Pardos, Baker, Pedro, Gowda, and Gowda (2014), within the web-based tutoring platform ASSISTments, used a complex set of parameters to identify students' confusion. Overall, the behaviours of confused students were characterized by (i) repeated sequential errors and (ii) incorrect skill-based actions prior to spending a long time on a current related task. Lee et al. (2011b) analyzed how novices learn to code. Students' confusion was found to be commonly associated with novices making repeated compilation errors. Similarly, while solving algebra problems, students who experienced confusion were those who had minimal correct responses and spent the longest time on problem solving (Lagud & Rodrigo, 2010). Another study that aimed to teach introductory computer

literacy with AutoTutor reported that when confused, participants seemed puzzled, they were unsure about how to continue, and they struggled to understand the material (Craig et al., 2004). Finally, it has been reported that students who experience a metacognitive mismatch (in this case confident but in error) would spend more time on the feedback (Anderson et al., 1971; Kulhavy, 1977; Kulhavy & Stock, 1989).

Given these findings, it is not unreasonable to expect that students who were confused during the *observe* phase of a POE task may spend longer on that task and may take several attempts on that task to correct their understanding. Moreover, unlike the previous studies, where students were provided with immediate corrective or confirmatory feedback, in the current simulation-based environment students were required to discover the correct responses through interactive engagement with the tasks. They could not proceed to the next task unless the current task was completed correctly. Therefore, students' efforts to find a correct response or to overcome their learning impasses could manifest in terms of their task attempts and time-on-task behaviours.

In this study, students' performance or conceptual understanding was measured using their scores from the *knowledge-transfer* task described above. Student groups were compared using the t-test, where the results are reported in terms of *p*-value statistic, *t*-value statistic, and Cohen's *d* for effect size. In the results section, the effect size can be interpreted as small when  $d = 0.20$ , medium when  $d = 0.50$ , and large when  $d = 0.80$  (Rosenthal, 1996; Rosenthal & Rosnow, 1984). The selected significance level was  $\alpha = 0.05$  (\*), and the tests were considered marginally significant when  $\alpha < 0.10$  (.). Because the analysis involved multiple comparisons, Benjamini–Hochberg (BH) post hoc correction was applied to control for false positives. This correction procedure has frequently been used in previous studies on students' task difficulties and their affective analyses (Botelho, Baker, Ocumpaugh, & Heffernan, 2018; Karumbaiah, Andres, Botelho, Baker, & Ocumpaugh, 2018; Nawaz, Srivastava, Yu, Baker, et al., 2020; Nawaz, Srivastava, Yu, Khan, et al., 2020; Ocumpaugh et al., 2017).

#### 4.4. Data Processing

In this study, data from the January 2016 offering of the course Habitable Worlds was analyzed. The trace data for this course consisted of 814,441 interaction entries, which were recorded in the system. Of these interactions, over 15,000 were related to the Stellar Lifecycles module.

One of the challenges in data processing was coming up with rules that could identify outliers for the session time. In a fully online environment, session time is most susceptible to errors (Sun, 2006). Therefore, as a first step, all module-based activities where students' session time exceeded five standard deviations (SD) from the median were excluded. This subjective judgment was made after analyzing the frequency tables, which revealed that nearly 99% of all module-based sessions were completed within this time. Other researchers have also used such procedures (del Valle & Duffy, 2009; Wise, Speer, Marbouti, & Hsiao, 2013). It was important to eliminate such interactions because it could be that students start a learning session and then leave the browser window open without actively engaging in a meaningful learning activity (e.g., in one instance a student's session time was 20 hours long).

Next, a cluster analysis was performed using module-based features such as mean module scores, mean module task completions, mean attempts, and mean time at the module tasks. The purpose of cluster analysis was to group the students based on the similarities of their behavioural patterns within the Stellar Lifecycles module. To determine the number of clusters on normalized data, a visual assessment of tendency (VAT) algorithm was run (Bezdek & Hathaway, 2002). Later, the k-means algorithm (MacQueen, 1967) was applied for clustering since it resulted in more balanced, well-connected, and well-separated clusters. To ensure the convergence of the centroid update, the algorithm was run multiple times (Kassambara, 2017). This analysis ultimately resulted in two clear groups or clusters of students (Group 1:  $n = 212$ ; Group 2:  $n = 130$ ). A comparison of the two groups seemed to suggest that

- students in both groups were persistent as they completed all task-related submissions;
- based on students' choice of hypothesis, both groups had similar levels of prior knowledge; a comparable proportion of students across the two groups proposed the correct hypothesis—34% of students from Group 1 and 32% from Group 2; and
- students' chosen hypotheses also revealed that they had misconceptions of a similar nature; the majority of students in both groups endorsed a common misconception—43% from Group 1 and 49% from Group 2.

Next, in accordance with existing works (Dienes, 2012; Nawaz, Kennedy, Bailey, Mead, & Horodyskyj, 2018), students' confidence in their proposed hypotheses was inferred through their free-text responses. During the prediction task, students were asked to report on why they proposed a hypothesis. The reasoning provided by students often included phrases such as "I am just guessing," "It's just a guess," "I am not sure, but this seems to be making sense." The authors labelled each response as guessing or not guessing behaviour and then attributed guessing to a "lack" of confidence (Dienes, 2012; Engelbrecht, Harding, & Potgieter, 2005). The labelling decision was based on students' expression of doubt regarding their selected hypotheses, or their reasoning for the selected hypotheses, or both. It was also found that students who were labelled as less confident generally used less technical language in the reasoning associated with hypotheses and they also wrote a shorter



amount of text while reasoning about the hypotheses (Nawaz et al., 2018). Overall, an analysis of the number of words entered by students, the number of scientific or key technical terms used by students, and students’ expression of guessing and doubts in their hypothesis-reasoning suggested that Group 1 students were more confident than the students in Group 2.

Regarding students’ confidence, Kang et al. (2011) suggest that there exists a link between confidence and the accuracy of students’ responses. Responses made with higher confidence are more likely to be accurate than responses made with lower confidence. Rangel, Möller, Sitter, Stibane, and Strzelczyk (2017) found that “confidence for incorrect answers is significantly lower than confidence for correct answers.” If this is the case, it would be expected that guessing is less likely to result in a correct response. When investigated in the current study, it was found that only 7% of students were guessing when they proposed a correct hypothesis—suggesting the validity of the operationalization of confidence.

Next, students’ task-based interactions during the observe phase were extracted, as this is where moments of confusion were expected. First, task-based interactions were extracted for students who make a correct prediction and those who make incorrect predictions. Then, these task-based interactions were refined (in terms of the response correctness) for the two student groups that had emerged from clustering. Last, the distribution of these task-based interactions was analyzed, and further cleaning was performed using time-oriented heuristics. In this case, an upper limit of 40 minutes was defined for time on the observe task. Studies have discussed the reliability of such measures (del Valle & Duffy, 2009; Kovanovic et al., 2015).

## 5. Results

In this paper, the goal was to identify specific moments of students’ confusion, as would be predicted by theory. Students’ confusion may be reflected in the way in which they respond to feedback, specifically in the *observation* phase of the POE task. To investigate generally whether students who make an incorrect prediction would be more likely to feel confused than students who make a correct prediction, comparisons between these students were made. Table 1 shows comparisons at a global level between students who made a correct prediction and those who made an incorrect prediction.

**Table 1:** Comparison of Students Who Made a Correct Prediction with Those Who Made Incorrect Predictions

	Global Comparison					
	Correct Prediction	Incorrect Prediction	<i>T</i>	<i>p</i>	<i>sig after BH correction</i>	<i>d</i>
	Mean ± SD	Mean ± SD				
Avg. Attempts	1.12 ± 0.46	1.34 ± 1.04	-2.74	<0.01	*	0.28
Avg. Time	3.88 ± 4.97	4.03 ± 3.80	-1.29	0.20		0.14

It can be seen from Table 1 that students who made incorrect predictions were likely to make significantly more attempts in the *observe* phase than students who made correct predictions, although the time the two groups spent on the task did not differ significantly.

The next set of analyses compared the behaviours of incorrectly predicting students: those who were high in confidence (Group 1) with those who were low in confidence (Group 2). Table 2 shows that there were no significant differences between students when it came to average task attempts. However, after making an incorrect prediction, Group 1 students spent marginally more time in the *observation* phase than Group 2 students.

**Table 2:** Comparison of the Incorrectly Predicting Students from Group 1 with the Incorrectly Predicting Students from Group 2

	Students Who Made <i>Incorrect</i> Predictions					
	Group 1—High Confidence	Group 2—Low Confidence	<i>T</i>	<i>p</i>	<i>sig after BH correction</i>	<i>d</i>
	Mean ± SD	Mean ± SD				
Avg. Attempts	1.25 ± 0.96	1.47 ± 1.15	-1.50	0.13		0.21
Avg. Time	4.40 ± 4.42	3.46 ± 2.50	2.03	<0.05	.	0.26

Table 3 presents the data only for high-confidence (Group 1) students. It compares those who made a correct prediction with those who made an incorrect prediction. Incorrectly predicting students in Group 1 were found to make significantly more attempts and spend a marginally longer time in the *observe* phase than students who made a correct prediction.

**Table 3:** Comparison of Group 1 Students Who Made Correct Predictions with Group 1 Students Who Made Incorrect Predictions

	Group 1—High Confidence		<i>T</i>	<i>p</i>	<i>sig after BH correction</i>	<i>d</i>
	Correct Prediction	Incorrect Prediction				
	Mean ± SD	Mean ± SD				
Avg. Attempts	1.01 ± 0.12	1.25 ± 0.96	-2.89	<0.00	*	0.35
Avg. Time	3.46 ± 2.26	4.40 ± 4.42	-2.02	<0.05	.	0.27

Finally, analyses were completed to determine whether those students who possibly experienced a *hypercorrection* effect or a metacognitive mismatch developed a better conceptual understanding than those students who did not. Previous research would suggest that a *hypercorrection* effect would be particularly apparent for a comparison of the incorrectly predicting students (comparison of high-confidence Group 1 students with low-confidence Group 2 students). Table 4 shows this comparison. While the results are marginally significant ( $\alpha < 0.1$ ), there is a clear trend in the expected direction (the incorrectly predicting students from Group 1 perform better; effect size = 0.24).

**Table 4:** Comparison of Incorrectly Predicting Students of Group 1 with Incorrectly Predicting Students of Group 2 at the *Knowledge-Transfer* Task

	Students Who Made <i>Incorrect</i> Predictions		<i>T</i>	<i>p</i>	<i>sig after BH correction</i>	<i>d</i>
	Group 1—High Confidence	Group 2—Low Confidence				
	Mean ± SD	Mean ± SD				
Transfer Score	7.62 ± 3.12	6.82 ± 3.74	1.67	0.06	.	0.24

To further investigate whether the *hypercorrection* effect influenced students’ conceptual understanding, the high-confidence (Group 1) correctly predicting students were compared with the high-confidence (Group 1) incorrectly predicting students at the *knowledge-transfer* task. Table 5 shows this comparison, and although no difference was found between the learning outcomes of these groups, it is, however, interesting to see that the high-confidence incorrectly predicting students could achieve similar learning outcomes as the high-confidence correctly predicting students.

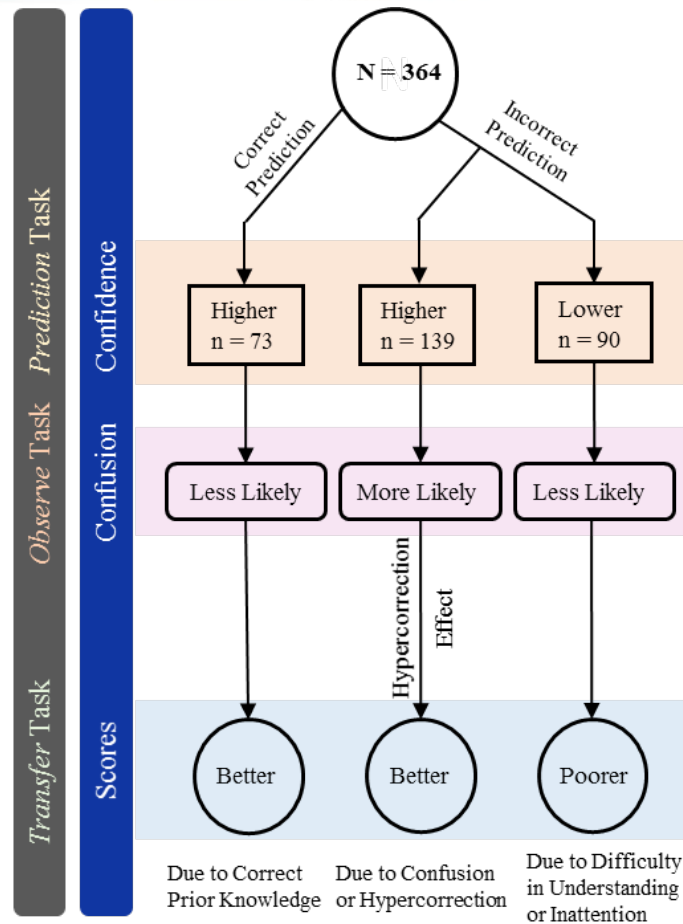
**Table 5:** Comparison of the Group 1 Incorrectly Predicting Students with the Group 1 Correctly Predicting Students at the *Knowledge-Transfer* Task

	Group 1—High Confidence		<i>T</i>	<i>p</i>	<i>sig after BH correction</i>	<i>d</i>
	Correct Prediction	Incorrect Prediction				
	Mean ± SD	Mean ± SD				
Transfer Score	7.04 ± 3.29	7.62 ± 3.12	-0.94	0.37		0.07

Most of the results presented above are in line with the theoretical predictions. However, it is important to note that the effect sizes associated with these findings are generally modest (Cohen, 1988). This is noted in the discussion section.

## 6. Discussion

Overall, this paper investigated whether specific moments of students’ confusion could be identified using behavioural markers derived from learning analytics in a digital, simulation-based learning environment. While, generally, students who made errors during their learning process were expected to show more signs of confusion, the concepts of *hypercorrection* and *metacognitive mismatch* suggest that students’ confidence plays a role in how they respond to errors (see Figure 4).



**Figure 4:** Moments of confusion in a POE-based simulation environment. Students’ confidence in their *predictions* may affect how they respond to errors during their *observations*. Students with higher confidence in their incorrect predictions are likely to be confused and *hypercorrected*, which may result in better learning outcomes.

Specifically, it was expected that students who were confident and made errors would show more signs of confusion than other students. The overall findings of this study indicate that students show signs of confusion in ways that align with theoretical predictions.

This study mainly investigated students who were high in confidence and made an incorrect prediction. These students, compared to those who made a high-confidence correct prediction, spent significantly more time and made marginally more attempts in the *observation* phase of the simulation-based learning task. They also spent marginally more time in this phase than students who were lower in confidence and made an incorrect prediction. These results are entirely consistent with, and predicted by, the *hypercorrection* effect, as described by Butterfield and Metcalfe (2001, 2006).

It is important to mention that contrary to the previous studies where students were provided with immediate corrective or confirmatory feedback (Butterfield & Metcalfe, 2001), students in the current simulation-based environment were required to discover the correct responses through interactive engagement with the tasks. They could not proceed to the next task unless the current task was completed correctly. Therefore, students’ efforts to find a correct response or to overcome their learning impasses could manifest in terms of their task attempts and time-on-task behaviours. In particular, their learning behaviours, such as making repeated attempts at the task and spending a greater amount of time on the task, are indicative of students who are attending to and responding to an error they have made and making an effort to solve or resolve their confusion about the concepts presented in the task. Previous studies that investigated students’ confusion have called these behaviours students’ “effort” (Lehman & Graesser, 2015) and “struggle” (Nawaz et al., 2018) toward confusion resolution.

With regard to the *hypercorrection* effect and students’ learning outcomes, Butterfield and Metcalfe (2001, 2006) suggest that students who experience a *hypercorrection* effect pay additional attention to the learning material, which results in their having a better understanding of the concepts they are covering. Before discussing the findings of this study, it is important to mention that the mark used for students’ assessment on the *knowledge-transfer* task was a more stringent measure than in the existing studies. Unlike most studies on the *hypercorrection* effect, where students were tested on general knowledge or non-

academic material (Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009, 2010; Metcalfe & Finn, 2011), this study tested them on academic content. Also, while the previous studies assessed students' performance on repeated MCQs (matching questions during the pre-test and post-test), this study measured students' performance on a knowledge-transfer task built on advanced material. It could, thus, be argued that the outcome measure in this study was more rigorous.

1. First, the *hypercorrection* effect would be borne out by higher-confidence students who made an incorrect prediction performing better (i.e., have better learning outcomes) than lower-confidence students who made a similar incorrect prediction. In the context of this investigation, although the difference between these groups was marginally significant ( $\alpha < 0.10$ ), the trends of the mean scores and the effect size were in line with theory.
2. Second, the *hypercorrection* effect could be observed when the learning outcomes of the high-confidence incorrectly predicting students are compared to the high-confidence correctly predicting students. While not better, the incorrectly predicting students achieved similar scores as the correctly predicting students. It could be that when the high-confidence incorrectly predicting students discovered an error, they were surprised and, thus, paid more attention. Apparently, it enabled them to achieve similar scores to those of the high-confidence correctly predicting students.

Because it was less clear from the literature whether students who make correct responses with lower confidence show any signs of confusion, this analysis was not presented in this paper. Taken together, the results tend to suggest that when students commit to a position within a simulation-based learning task with some confidence and then find they have made a mistake, they get a bit confused and subsequently pay more attention. As a result, they put in additional effort, which may ultimately lead them to form a better understanding. This is entirely consistent with the cognitive conflict (Limón, 2001), cognitive disequilibrium (Festinger, 1957; Graesser et al., 2005; Piaget, 1952a), and impasse-driven theories of learning (Brown & VanLehn, 1980; VanLehn et al., 2003b), suggesting that effortful cognitive activities can promote confusion resolution and ultimately benefit student learning. The findings of this paper are also compatible with previous empirical studies on confusion (Lee et al., 2011; Liu, Pataranutaporn, Ocumpaugh, & Baker, 2013).

Chen, Pan, Sung, & Chang (2013) report that students' erroneous predictions in POE-based learning designs can result in cognitive conflict, dissonance, or disequilibrium, which can ultimately lead to students' dissatisfaction with their current views and may help them bring about a conceptual change and remediation of misconceptions. Overall, while the previous studies suggest that POE environments can probe students' prior knowledge (Liew & Treagust, 1995) and can promote a conceptual change (Chairam, Somsook, & Coll, 2009), the current study seems to suggest that students' confidence in prior knowledge can affect their behaviours relating to error correction. In particular, this study proposes that students who face a metacognitive mismatch and associated hypercorrection are more likely to benefit from the POE instructional design than students who have lower confidence in their prior knowledge.

Moreover, some of the prior works that analyzed students' confusion experimentally manipulated cognitive disequilibrium and reported that "cognitive disequilibrium causes an increase in uncertainty and allegedly confusion" (D'Mello et al., 2014; Lehman, D'Mello, & Graesser, 2012; Lehman et al., 2011). This study showed that a naturalistic simulation-based POE learning environment has the potential to naturally trigger a cognitive conflict or disequilibrium in students, especially for those students whose expectations do not match their predictions (Fensham & Kass, 1988). The present study also suggests that POE environments might represent a good candidate for learning design for future research on emotions and affect.

## 7. Conclusion

Overall, this study investigated students' *moments* of confusion in line with theory. The value of having a strong theory to aid in the interpretation of the learning analytic measures used in this study should not be underestimated. Several researchers have argued the critical need to connect analytics with learning theory for further advancement in research and practice of learning analytics. In one study, while examining the association between students' trace data and their learning outcomes, the importance of instructional conditions was empirically shown (Gašević, Dawson, Rogers, & Gasevic, 2016). Researchers have further suggested that to gain "actionable" insights into students' learning progress, to improve study designs, and to enhance and interpret the study findings, existing theory should be utilized (Gašević, Dawson, & Siemens, 2014).

This study presented an innovative methodological approach to confusion research. Through cluster analysis, two groups of students were identified who, broadly speaking, exhibited similar levels of persistence (because they completed all the module tasks) and had similar levels of prior knowledge and misconceptions (based on their selection of hypotheses at the prediction task). The only apparent difference between the two groups was their confidence in prior knowledge—which was inferred from students' free-text data that they wrote while providing the reasoning for their choice of hypotheses. This difference in students' confidence seemed to influence how they interacted within a simulation-based POE environment—in particular, during the observation phase, which was the focus of this study.

There does seem to be value in looking for digital markers of confusion, particularly during the observe phase of a simulation-based POE environment. Broadly speaking, knowledge of student profiles in terms of their confidence and prior



knowledge can help interpret their effort associated with confusion. This can ultimately enable timely interventions. Interventions or feedback is important in learning (Shute, 2008); feedback is “the process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work itself, in order to generate improved work” (Boud & Molloy, 2012). The importance of interventions in learning analytics to “close the feedback loop” has been emphasized by several researchers (Barker & Pinard, 2014; Clow, 2012, 2014). Within the current context, student intervention can prevent persistent or prolonged confusion, which has been associated with poor learning outcomes and performance (D’Mello et al., 2009; Lee et al., 2011).

Knowing when students get confused or which students are more likely to be confused could be useful for the provision of automated feedback by using learning analytic techniques. On the one hand, this can allow educators to identify students who need more guidance and support. This approach, where teachers offer dynamically planned interventions to one or more students, is termed “proactive remediation” (Miller et al., 2015). Through the teacher’s assistance, this can possibly lead to students’ enhanced engagement and persistence in the task. On the other hand, a knowledge of the likely moments of students’ confusion could help educators know when the material is getting more complex, so that they may change the pace of the task or include additional support.

The approach to confusion detection used in this research leverages learning analytics. The only data used in this study is students’ interaction-based trace data. There is potential, therefore, to deploy the processes and findings from this study at scale. In time, these measures may be used as the basis for digital systems that can automatically detect and respond to learners’ difficulties and cater to their confusion. A key challenge with such a development would be knowing the precise moment to provide feedback to confused students. Immediate feedback could hinder the productive educational benefits of being confused, while overly delayed feedback may risk students disengaging from the task (Graesser & D’Mello, 2012).

## 8. Limitations and Future Work

There are some limitations to the present analysis that need to be addressed in the future. This preliminary study only included the *observe* phase of the POE learning design, given that this is where students’ *moments* of confusion were expected. In future, students’ emotions in the complete POE cycle could be analyzed. This would allow a more in depth understanding of the various emotions that students can undergo in a complete POE-based design. For example, future research could use learning analytic techniques to explore the expectation that students with sustained misconceptions from the *observation* task would likely be frustrated as they arrive at the *explanation* tasks.

In future, this study could also benefit by including external criterion, such as students’ self-reports. Researchers in learning sciences have used self-report instruments extensively (Colthorpe, Zimbardi, Ainscough, & Anderson, 2015; Ellis, Han, & Pardo, 2017; Gašević, Jovanovic, Pardo, & Dawson, 2017; Rodriguez et al., 2019). However, self-reports pose some challenges of their own, such as being influenced by social desirability bias (Krosnick, 1999; Paulhaus, 1991). This refers to the inclination of people to answer in line with what society or researchers view as favourable rather than their actual behaviours, attitudes, and beliefs. Another issue with self-report is that people often have a different conceptual understanding of the questions, and hence the participants may interpret the self-report items differently (White, 1989). Moreover, it has been suggested that self-reports can influence cognition or affects in subtle ways (Lazarus, 1991). Additionally, trace data and self-reports on the same construct often lack an association (Winne & Jamieson-Noel, 2002), and learners’ poor reflection is a likely reason for this mismatch (Zhou & Winne, 2012). Last, in some of the prior studies, trace data has been found to have a stronger association with students’ learning outcomes and study tactics than their self-report measures (Winne & Jamieson-Noel, 2002; Zhou & Winne, 2012).

Another possible future direction within a POE framework could be to explore how students select a hypothesis and then gather information about it. For this, students could be asked to express their confidence in each of the options with which they are presented. Research suggests that given a complex learning task, the participants can test their hypotheses through either selection or reception (Markant & Gureckis, 2014b). The different approaches participants adopt can be affected by their level of uncertainty or confidence, which, in turn, can influence their learning (Markant & Gureckis, 2014a; Markant et al., 2016).

## Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

## Funding

Sadia Nawaz is supported by the Research Training Program (RTP) Scholarship, the Melbourne Research Scholarship, and the Science of Learning Research Center (SLRC) top-up scholarship. Any opinions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

## Acknowledgements

We are thankful to the anonymous reviewers for their time and valuable feedback on this paper. The quality of this manuscript has improved because of their insightful suggestions. We also wish to thank Namrata Srivastava for her feedback on this work.

## References

- Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology*, 62(2), 148–156. <https://dx.doi.org/10.1037/h0030766>
- Arguel, A., & Lane, R. (2015). Fostering deep understanding in geography by inducing and managing confusion: an online learning approach. In T. Reiners et al. (Eds.), *Globally connected, digitally enabled. Proceedings of the 32nd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2015)*, 29 November–2 December 2015, Perth, Western Australia (pp. 373–377). Australasian Society for Computers in Learning in Tertiary Education. Retrieved from <http://www.2015conference.ascilite.org/wp-content/uploads/2015/11/ascilite-2015-proceedings.pdf>
- Arguel, A., Lockyer, L., Kennedy, G., Lodge, J., & Pachman, M. (2019). Seeking optimal confusion: A review on epistemic emotion management in interactive digital learning environments. *Interactive Learning Environments*, 27(2), 200–210. <https://dx.doi.org/10.1080/10494820.2018.1457544>
- Arguel, A., Lockyer, L., Lipp, O. V., Lodge, J. M., & Kennedy, G. (2017). Inside out: Detecting learners' confusion to improve interactive digital learning environments. *Journal of Educational Computing Research*, 55(4), 526–551. <https://dx.doi.org/10.1177%2F0735633116674732>
- Baker, R., D'Mello, S., Rodrigo, M. M. T., & Graesser, A. (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. <https://dx.doi.org/10.1016/j.ijhcs.2009.12.003>
- Baker, R., Kalka, J., Alevan, V., Rossi, L., Gowda, S. M., Wagner, A. Z., ..., Ocumpaugh, J. (2012). Towards sensor-free affect detection in cognitive tutor algebra. In K. Yacef et al. (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining (EDM2012)*, 19–21 June 2012, Chania, Greece (pp. 126–133). International Educational Data Mining Society. Retrieved from [https://educationaldatamining.org/EDM2012/uploads/procs/EDM\\_2012\\_proceedings.pdf](https://educationaldatamining.org/EDM2012/uploads/procs/EDM_2012_proceedings.pdf)
- Baker, R., Moore, G. R., Wagner, A. Z., Kalka, J., Salvi, A., Karabinos, M., & Yaron, D. (2011). The dynamics between student affect and behavior occurring outside of educational software. In *Proceedings of the 2011 Conference on Affective Computing and Intelligent Interaction (ACII '11)*, 9–12 October 2011, Memphis, TN, USA (pp. 14–24). Springer. [https://dx.doi.org/10.1007/978-3-642-24600-5\\_5](https://dx.doi.org/10.1007/978-3-642-24600-5_5)
- Ballantyne, R., & Bain, J. (1995). Enhancing environmental conceptions: An evaluation of cognitive conflict and structured controversy learning units. *Studies in Higher Education*, 20(3), 293–303. <https://dx.doi.org/10.1080/03075079512331381565>
- Barker, M., & Pinard, M. (2014). Closing the feedback loop? Iterative feedback between tutor and student in coursework assessments. *Assessment & Evaluation in Higher Education*, 39(8), 899–915. <https://dx.doi.org/10.1080/02602938.2013.875985>
- Bezdek, J. C., & Hathaway, R. J. (2002). VAT: A tool for visual assessment of (cluster) tendency. In *Proceedings of the 2002 International Joint Conference on Neural Networks (IJCNN 2002)*, 12–17 May 2002, Honolulu, HI, USA (pp. 2225–2230). IEEE Computer Society. <https://dx.doi.org/10.1109/IJCNN.2002.1007487>
- Bjork, R. A. (2011). *Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning*. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, J. R. Pomerantz (Eds.), & FABBS Foundation, *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers. Retrieved from <https://psycnet.apa.org/record/2011-19926-008>
- Bosch, N., D'Mello, S., & Mills, C. (2013). What emotions do novices experience during their first computer programming learning session? In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED '13)*, 9–13 July 2013, Memphis, TN, USA (pp. 11–20). Springer. [https://dx.doi.org/10.1007/978-3-642-39112-5\\_2](https://dx.doi.org/10.1007/978-3-642-39112-5_2)
- Botelho, A. F., Baker, R., Ocumpaugh, J., & Heffernan, N. T. (2018). Studying affect dynamics and chronometry using sensor-free detectors. In K. E. Boyer & M. Yudelson (Eds.), *Proceedings of the 11th International Conference on Educational Data Mining (EDM2018)*, 15–18 July 2018, Buffalo, NY, USA (pp. 157–166). International Educational Data Mining Society. Retrieved from [https://educationaldatamining.org/files/conferences/EDM2018/EDM2018\\_Preface\\_TOC\\_Proceedings.pdf](https://educationaldatamining.org/files/conferences/EDM2018/EDM2018_Preface_TOC_Proceedings.pdf)

- Boud, D., & Molloy, E. (2012). *Feedback in higher and professional education: Understanding it and doing it well* (1st Ed.); London, UK: Routledge. Retrieved from <https://www.routledge.com/Feedback-in-Higher-and-Professional-Education-Understanding-it-and-doing/Boud-Molloy/p/book/9780415692298> <https://dx.doi.org/10.4324/9780203074336>
- Bowen, M., & Haysom, J. (2014). *Predict, observe, explain: Activities enhancing science understanding*. Arlington, VA: NSTA Press. Retrieved from <https://my.nsta.org/resource/2625>
- Brown, J., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. In *Cognitive Science*, 4(4), 379–426. [https://dx.doi.org/10.1016/S0364-0213\(80\)80010-3](https://dx.doi.org/10.1016/S0364-0213(80)80010-3)
- Bunderson, V. C., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (Vol. 3). New York, NY: Macmillan. <https://dx.doi.org/10.1002/j.2330-8516.1988.tb00291.x>
- Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 918–928. <https://dx.doi.org/10.1037/0278-7393.34.4.918>
- Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1491–1494. <https://dx.doi.org/10.1037/0278-7393.27.6.1491>
- Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, 1(1), 69–84. <https://dx.doi.org/10.1007/s11409-006-6894-z>
- Chairam, S., Somsook, E., & Coll, R. K. (2009). Enhancing Thai students' learning of chemical kinetics. *Research in Science & Technological Education*, 27(1), 95–115. <https://dx.doi.org/10.1080/02635140802658933>
- Chen, Y. L., Pan, P. R., Sung, Y. T., & Chang, K. E. (2013). Correcting misconceptions on electronics: Effects of a simulation-based learning environment backed by a conceptual change model. *Journal of Educational Technology & Society*, 16(2), 212–227. Retrieved from <https://drive.google.com/file/d/1dgiywBcc1xq6fKyGmHUzXw6DzxsD2Z1/view>
- Chiu, C. Y., Hong, Y. Y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology*, 73, 19–30. <https://dx.doi.org/10.1037/0022-3514.73.1.19>
- Clore, G. L. (1992). Cognitive phenomenology: Feelings and the construction of judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 133–163). Hillsdale, NJ: Lawrence Erlbaum Associates. Retrieved from <https://psycnet.apa.org/record/1992-98414-005>
- Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In S. Buckingham Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*, 29 April–2 May 2012, Vancouver, BC, Canada (pp. 134–138). New York, NY: ACM. <https://dx.doi.org/10.1145/2330601.2330636>
- Clow, D. (2014). Data wranglers: Human interpreters to help close the feedback loop. In *Proceedings of the 4th International Conference on Learning Analytics and Knowledge (LAK '14)*, 24–28 March 2014, Indianapolis, IN, USA (pp. 49–53). New York, NY: ACM. <https://dx.doi.org/10.1145/2567574.2567603>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ, USA: Lawrence Erlbaum Associates. <https://dx.doi.org/10.4324/9780203771587>
- Colthorpe, K., Zimbardi, K., Ainscough, L., & Anderson, S. (2015). Know thy student! Combining learning analytics and critical reflections to increase understanding of students' self-regulated learning in an authentic setting. *Journal of Learning Analytics*, 2(1), 134–155. <https://dx.doi.org/10.18608/jla.2015.21.7>
- 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. <https://dx.doi.org/10.1080/1358165042000283101>
- Dalziel, J. (2010). *Practical eTeaching strategies for predict—observe—explain, problem-based learning and role plays*. Sydney, Australia: LAMS International. Retrieve from <https://practicaleteachingstrategies.com/files/LAMSbook2010.Final.pdf>
- del Valle, R., & Duffy, T. M. (2009). Online learning: Learner characteristics and their approaches to managing learning. *Instructional Science*, 37(2), 129–149. <https://dx.doi.org/10.1007/s11251-007-9039-0>
- Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition*, Vol. 1 (pp. 337–364). Berlin, Germany: De Gruyter Mouton. <https://dx.doi.org/10.1515/9781934078242.337>
- D'Mello, S., Craig, S. D., Gholson, B., Franklin, S., Picard, R., & Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. In *Affective interactions: The computer in the affective loop workshop at 2005 International Conference on Intelligent User Interfaces* (pp. 7–13). New York, NY: ACM Press.



- D'Mello, S., & Graesser, A. (2010). *Modeling cognitive-affective dynamics with hidden Markov models*. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 11–14 August 2010, Portland, OR, USA (pp. 2721–2726). Austin, TX: Cognitive Science Society. Retrieved from [https://cognitivesciencesociety.org/wp-content/uploads/2019/01/cogsci10\\_proceedings.pdf](https://cognitivesciencesociety.org/wp-content/uploads/2019/01/cogsci10_proceedings.pdf)
- D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition and Emotion*, 25(7), 1299–1308. <https://dx.doi.org/10.1080/02699931.2011.613668>
- D'Mello, S., & Graesser, A. (2012a). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157. <https://dx.doi.org/10.1016/j.learninstruc.2011.10.001>
- D'Mello, S., & Graesser, A. (2012b). Emotions during learning with AutoTutor. In P. Durlach & A. Lesgold (Eds.), *Adaptive technologies for training and education* (pp. 117–139). Cambridge, UK: Cambridge University Press. <https://dx.doi.org/10.1017/CBO9781139049580.010>
- D'Mello, S., & Graesser, A. (2014a). Confusion. In R. Pekrun & L. G. Lisa (Eds.), *International handbook of emotions in education* (pp. 289–310). Hoboken, NJ, USA: Taylor and Francis.
- D'Mello, S., & Graesser, A. (2014b). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychologica*, 151, 106–116. <https://dx.doi.org/10.1016/j.actpsy.2014.06.005>
- D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153–170. <https://dx.doi.org/10.1016/j.learninstruc.2012.05.003>
- D'Mello, S., Person, N., & Lehman, B. (2009). Antecedent-consequent relationships and cyclical patterns between affective states and problem solving outcomes. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. Graesser (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED '09)*, 6–10 July 2009, Brighton, UK (pp. 57–64). IOS Press. <https://dl.acm.org/doi/10.5555/1659450.1659463>
- D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, 22(4), 53–61. <https://dx.doi.org/10.1109/MIS.2007.79>
- D'Mello, S., Taylor, R. S., & Graesser, A. (2007). Monitoring affective trajectories during complex learning. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society (CogSci 2007)*, 1–4 August 2007, Nashville, TN, USA (pp. 197–202). Austin, TX: Cognitive Science Society.
- Driver, R. (1983). *The pupil as scientist?* Milton Keynes, UK: Open University Press.
- Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671–688. <https://dx.doi.org/10.1080/09500690305016>
- Ellis, R. A., Han, F., & Pardo, A., (2017). Improving learning analytics—Combining observational and self-report data on student learning. *Educational Technology & Society*, 20(3), 158–169. Retrieved from <https://www.jstor.org/stable/26196127>
- Engelbrecht, J., Harding, A., & Potgieter, M. (2005). Undergraduate students' performance and confidence in procedural and conceptual mathematics. *International Journal of Mathematical Education in Science and Technology*, 36(7), 701–712. <https://dx.doi.org/10.1080/00207390500271107>
- Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, 16(1), 88–92. <https://dx.doi.org/10.3758/PBR.16.1.88>
- Fazio, L. K., & Marsh, E. J. (2010). Correcting false memories. *Psychological Science*, 21(6), 801–803. <https://dx.doi.org/10.1177/0956797610371341>
- Fensham, P. J., & Kass, H. (1988). Inconsistent or discrepant events in science instruction. *Studies in Science Education*, 15(1), 1–16. <https://dx.doi.org/10.1080/03057268808559946>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press. Retrieved from <https://psycnet.apa.org/record/1993-97948-000>
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting learning success. *The Internet and Higher Education*, 28, 68–84. <https://dx.doi.org/10.1016/j.iheduc.2015.10.002>
- Gašević, D., Dawson, S., & Siemens, G. (2014). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://dx.doi.org/10.1007/s11528-014-0822-x>
- Gašević, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, 4(2), 113–128. <https://dx.doi.org/10.18608/jla.2017.42.10>
- Graesser, A. (2011). Learning, thinking, and emoting with discourse technologies. *American Psychologist*, 66(8), 746–757. <https://dx.doi.org/10.1037/a0024974>
- Graesser, A., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology*, 10(7), 17–31. [https://dx.doi.org/10.1002/\(SICI\)1099-0720\(199611\)10:7<17::AID-ACP435>3.0.CO;2-7](https://dx.doi.org/10.1002/(SICI)1099-0720(199611)10:7<17::AID-ACP435>3.0.CO;2-7)



- Graesser, A., & D'Mello, S. (2012). Emotions during the learning of difficult material. *Psychology of Learning and Motivation*, 57, 183–225. <https://dx.doi.org/10.1016/B978-0-12-394293-7.00005-4>
- Graesser, A., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory & Cognition*, 33(7), 1235–1247. <https://dx.doi.org/10.3758/BF03193225>
- Graesser, A., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95(3), 524–536. <https://dx.doi.org/10.1037/0022-0663.95.3.524>
- Graesser, A., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 112–141). New York, NY: Guilford. Retrieved from <https://psycnet.apa.org/record/2010-03564-007>
- Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, 13, 74–83. <https://dx.doi.org/10.1016/j.edurev.2014.10.002>
- Gunstone, R. F. (1990). Children's science: A decade of developments in constructivist views of science teaching and learning. *The Australian Science Teachers Journal*, 36(4).
- Gunstone, R. F., & White, R. T. (1980). A matter of gravity. *Research in Science Education*, 10(1), 35–44. <https://dx.doi.org/10.1007/BF02356307>
- Horodyskyj, L. B., Mead, C., Belinson, Z., Buxner, S., Semken, S., & Anbar, A. D. (2018). Habitable Worlds: Delivering on the promises of online education. *Astrobiology*, 18(1), 86–99. <https://dx.doi.org/10.1089/ast.2016.1550>
- Huang, H. M. (2002). Toward constructivism for adult learners in online learning environments. *British Journal of Educational Technology*, 33(1), 27–37. <https://dx.doi.org/10.1111/1467-8535.00236>
- Kang, S. H., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, 103(1), 48–59. <https://dx.doi.org/10.1037/a0021977>
- Kapur, M. (2016). Examining productive failure, productive success, unproductive failure and unproductive success in learning. *Educational Psychologist*, 51(2), 289–299. <https://dx.doi.org/10.1080/00461520.2016.1155457>
- Karamustafaoğlu, S., & Mamlok-Naaman, R. (2015). Understanding electrochemistry concepts using the predict-observe-explain strategy. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 923–936. <https://dx.doi.org/10.12973/eurasia.2015.1364a>
- Karumbaiah, S., Andres, J. M. A. L., Botelho, A. F., Baker, R., & Ocumpaugh, J. (2018). The implications of a subtle difference in the calculation of affect dynamics. In *Proceedings of the 26th International Conference on Computers in Education (ICCE '18)*, 26–30 November 2018, Manila, Philippines (pp. 29–38). Philippines: Asia-Pacific Society for Computers in Education. Retrieved from <http://icce2018.ateneo.edu/wp-content/uploads/2018/12/C1-04.pdf>
- Kassambara, A. (2017). *Practical guide to cluster analysis in R—Unsupervised machine learning*. STHDA. Retrieved from <https://www.sthda.com>
- Keltner, D., & Shiota, M. N. (2003). New displays and new emotions: A commentary on Rozin and Cohen (2003). *Emotion*, 3, 86–91. <https://dx.doi.org/10.1037/1528-3542.3.1.86>
- Kendeou, P., & Broek, P. V. D. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, 97(2). <https://dx.doi.org/10.1037/0022-0663.97.2.235>
- Kennedy, G., & Lodge, J. (2016). All roads lead to Rome: Tracking students' affect as they overcome misconceptions. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), *Show Me The Learning: Proceedings of the 33rd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2016)*. 28–30 November 2016, Wellington, New Zealand (pp. 318–328). Australasian Society for Computers in Learning in Tertiary Education. Retrieved from [https://2016conference.ascilite.org/wp-content/uploads/ascilite2016\\_kennedy\\_full.pdf](https://2016conference.ascilite.org/wp-content/uploads/ascilite2016_kennedy_full.pdf)
- Kibirige, I., Osodo, J., & Tlala, K. M. (2014). The effect of predict-observe-explain strategy on learners' misconceptions about dissolved salts. *Mediterranean Journal of Social Sciences*, 5(4), 300–310. <https://dx.doi.org/10.5901/mjss.2014.v5n4p300>
- Kostyuk, V., Almeda, M., & Baker, R. (2018). Correlating affect and behavior in reasoning mind with state test achievement. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*, 5–9 March 2018, Sydney, Australia (pp. 26–30). New York, NY: ACM. <https://dx.doi.org/10.1145/3170358.3170378>
- Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., Baker, R., & Hatala, M. (2015). Does time-on-task estimation matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, 2(3), 81–110. <https://dx.doi.org/10.18608/jla.2015.23.6>
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50, 537–567. <https://dx.doi.org/10.1146/annurev.psych.50.1.537>

- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47(2), 211–232. <https://dx.doi.org/10.3102/00346543047002211>
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1(4), 279–308. <https://dx.doi.org/10.1007/BF01320096>
- Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology*, 68(5), 522–528. <https://dx.doi.org/10.1037/0022-0663.68.5.522>
- Lagud, M. C. V., & Rodrigo, M. M. T. (2010). The affective and learning profiles of students using an intelligent tutoring system for algebra. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS 2010)*, 14–18 June 2010, Pittsburgh, PA, USA, (pp. 255–263). Berlin, Germany: Springer. [https://dx.doi.org/10.1007/978-3-642-13388-6\\_30](https://dx.doi.org/10.1007/978-3-642-13388-6_30)
- Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, 46, 352–367. <https://dx.doi.org/10.1037/0003-066X.46.4.352>
- Lee, D. M. C., Rodrigo, M. M. T., Baker, R., Sugay, J. O., & Coronel, A. (2011). Exploring the relationship between novice programmer confusion and achievement. In S. K. D’Mello, A. C. Graesser, B. W. Schuller, & Jean-Claude Martin (Eds.), *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII ’11)*, 9–12 October 2011, Memphis, TN, USA (pp. 175–184). Berlin, Germany: Springer. [https://dx.doi.org/10.1007/978-3-642-24600-5\\_21](https://dx.doi.org/10.1007/978-3-642-24600-5_21)
- Lehman, B., D’Mello, S., Chauncey, A., Gross, M., Dobbins, A., Wallace, P., ..., Graesser, A. (2011). Inducing and tracking confusion with contradictions during critical thinking and scientific reasoning. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education (AIED ’11)*, 28 June–2 July 2011, Auckland, New Zealand (pp. 171–178). Berlin, Germany: Springer. [https://dx.doi.org/10.1007/978-3-642-21869-9\\_24](https://dx.doi.org/10.1007/978-3-642-21869-9_24)
- 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. <https://dx.doi.org/10.1016/j.iheduc.2012.01.002>
- Lehman, B., & Graesser, A. (2015). To resolve or not to resolve? That is the big question about confusion. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Proceedings of the 17th International Conference on Artificial Intelligence in Education (AIED ’15)*, 22–26 June 2015, Madrid, Spain (pp. 216–225). Berlin, Germany: Springer. [https://dx.doi.org/10.1007/978-3-319-19773-9\\_22](https://dx.doi.org/10.1007/978-3-319-19773-9_22)
- Liew, C. W., & Treagust, D. F. (1995). A predict-observe-explain teaching sequence for learning about students’ understanding of heat and expansion of liquids. *Australian Science Teachers’ Journal*, 41(1), 68–71. Retrieved from [https://www.researchgate.net/publication/234752631\\_A\\_Predict-Observe-Explain\\_Teaching\\_Sequence\\_for\\_Learning\\_about\\_Students'\\_Understanding\\_of\\_Heat\\_and\\_Expansion\\_Liquids](https://www.researchgate.net/publication/234752631_A_Predict-Observe-Explain_Teaching_Sequence_for_Learning_about_Students'_Understanding_of_Heat_and_Expansion_Liquids)
- Liew, C. W., & Treagust, D. F. (1998). *The effectiveness of predict-observe-explain tasks in diagnosing students’ understanding of science and in identifying their levels of achievement*. Retrieved from <https://eric.ed.gov/?id=ED420715>
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, 11(4-5), 357–380. [https://dx.doi.org/10.1016/S0959-4752\(00\)00037-2](https://dx.doi.org/10.1016/S0959-4752(00)00037-2)
- Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. (2013). Sequences of frustration and confusion, and learning. In S. K. D’Mello et al. (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining (EDM2013)*, 6–9 July 2013, Memphis, TN, USA (pp. 114–120). International Educational Data Mining Society/Springer. Retrieved from <https://files.eric.ed.gov/fulltext/ED558215.pdf>
- Lodge, J., Kennedy, G., Lockyer, L., Arguel, A., & Pachman, M. (2018). Understanding difficulties and resulting confusion in learning: An integrative review. *Frontiers in Education*, 3, 49. <https://dx.doi.org/10.3389/educ.2018.00049>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1 January 1967, Oakland, CA, USA (pp. 281–297).
- Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. New York, NY: Norton.
- Mansour, B. E., & Mupinga, D. M. (2007). Students’ positive and negative experiences in hybrid and online classes. *College Student Journal*, 41(1). Retrieved from <https://eric.ed.gov/?id=EJ765422>
- Markant, D., & Gureckis, T. M. (2014a). A preference for the unpredictable over the informative during self-directed learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society (CogSci 2014)*, 23–26 July 2014, Quebec City, QC, Canada (pp. 958–963). Austin, TX: Cognitive Science Society. Retrieved from <https://cogsci.mindmodeling.org/2014/papers/172/paper172.pdf>
- Markant, D. B., & Gureckis, T. M. (2014b). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, 143(1), 94–122. <https://dx.doi.org/10.1037/a0032108>

- Markant, D. B., Settles, B., & Gureckis, T. M. (2016). Self-directed learning favors local, rather than global, uncertainty. *Cognitive Science*, 40(1), 100–120. <https://dx.doi.org/10.1111/cogs.12220>
- McLeod, D. B. (1989). The role of affect in mathematical problem solving. In D. B. McLeod & V. M. Adams (Eds.), *Affect and Mathematical Problem Solving* (pp. 20–36). New York, NY: Springer. [https://dx.doi.org/10.1007/978-1-4612-3614-6\\_2](https://dx.doi.org/10.1007/978-1-4612-3614-6_2)
- McQuiggan, S. W., Robison, J. L., & Lester, J. C. (2010). Affective transitions in narrative-centered learning environments. *Educational Technology & Society*, 13(1), 40–53. Retrieved from <https://www.jstor.org/stable/jeductechsoci.13.1.40>
- Metcalfe, J., & Finn, B. (2011). People’s hypercorrection of high-confidence errors: Did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2). <https://dx.doi.org/10.1037/a0021962>
- Miller, W. L., Baker, R., Labrum, M. J., Petsche, K., Liu, Y. H., & Wagner, A. Z. (2015). Automated detection of proactive remediation by teachers in Reasoning Mind classrooms. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge (LAK '15)*, 16–20 March 2015, Poughkeepsie, NY, USA (pp. 290–294). New York, NY: ACM. <https://dx.doi.org/10.1145/2723576.2723607>
- Nawaz, S., Kennedy, G., Bailey, J., Mead, C., & Horodyskyj, L. (2018). Struggle town? Developing profiles of student confusion in simulation-based learning environments. *Proceedings of the 35th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2018)*, 25–28 November 2018, Geelong, Australia (pp. 224–233). Australasian Society for Computers in Learning in Tertiary Education. Retrieved from <https://2018conference.ascilite.org/wp-content/uploads/2018/12/ASCILITE-2018-Proceedings-Final.pdf>
- Nawaz, S., Srivastava, N., Yu, J. H., Baker, R., Kennedy, G., & Bailey, J. (2020). Analysis of task difficulty sequences in a simulation-based POE environment. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Proceedings of the 21st International Conference on Artificial Intelligence in Education (AIED '20)*, 6–10 July 2020, Ifrane, Morocco (virtual conference due to COVID-19) (pp. 423–436). Berlin, Germany: Springer. [https://dx.doi.org/10.1007/978-3-030-52237-7\\_34](https://dx.doi.org/10.1007/978-3-030-52237-7_34)
- Nawaz, S., Srivastava, N., Yu, J. H., Khan, A. A., Kennedy, G., Bailey, J., & Baker, R. (2020). How difficult is the task for you? Modelling and analysis of students’ task difficulty sequences. *International Journal of Artificial Intelligence in Education (IJAIED)*, submitted.
- Ocuppaugh, J., Andres, J. M., Baker, R., DeFalco, J., Paquette, L., Rowe, J., ... , Sottolare, R. (2017). Affect dynamics in military trainees using vMedic: From engaged concentration to boredom to confusion. In E. André, R. S. Baker, X. Hu, M. M. T. Rodrigo, & B. du Boulay (Eds.), *Proceedings of the 18th International Conference on Artificial Intelligence in Education (AIED 2017)*, 28 June–1 July 2017, Wuhan, China (pp. 238–249). Berlin, Germany: Springer. [https://dx.doi.org/10.1007/978-3-319-61425-0\\_20](https://dx.doi.org/10.1007/978-3-319-61425-0_20)
- Pachman, M., Arguel, A., & Lockyer, L. (2015). Learners’ confusion: Faulty prior knowledge or a metacognitive monitoring error? In T. Reiners et al. (Eds.), *Globally connected, digitally enabled. Proceedings of the 32nd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2015)*, 29 November–2 December 2015, Perth, Western Australia (pp. 522–526). Australasian Society for Computers in Learning in Tertiary Education. Retrieved from <http://www.2015conference.ascilite.org/wp-content/uploads/2015/11/ascilite-2015-proceedings.pdf>
- Pardos, Z. A., Baker, R. S. J. D., Pedro, M. S., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Journal of Learning Analytics*, 1(1), 107–128. <https://dx.doi.org/10.18608/jla.2014.11.6>
- Pardos, Z. A., & Horodyskyj, L. (2019). Analysis of student behaviour in *Habitable Worlds* using continuous representation visualization. *Journal of Learning Analytics*, 6(1), 1–15. <https://dx.doi.org/10.18608/jla.2019.61.1>
- Paulhaus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes, Vol. 1. Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press. <https://dx.doi.org/10.1016/B978-0-12-590241-0.50006-X>
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students’ self-regulated learning and achievement: A program of quantitative and qualitative research. *Educational Psychologist*, 37(2), 91–106. [https://dx.doi.org/10.1207/S15326985EP3702\\_4](https://dx.doi.org/10.1207/S15326985EP3702_4)
- Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urda, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA handbooks in psychology®. APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (pp. 3–31). Washington, DC: American Psychological Association. <https://dx.doi.org/10.1037/13274-001>
- Piaget, J. (1952a). *The origins of intelligence*. New York, NY: International University Press.
- Piaget, J. (1952b). *The origins of intelligence in children* (Vol. 8). New York, NY: International University Press.



- Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago, IL: University of Chicago Press.
- Rangel, R. H., Möller, L., Sitter, H., Stibane, T., & Strzelczyk, A. (2017). Sure, or unsure? Measuring students' confidence and the potential impact on patient safety in multiple-choice questions. *Medical Teacher*, 39(11), 1189–1194. <https://dx.doi.org/10.1080/0142159X.2017.1362103>
- Rodrigo, M. M. T., Baker, R. S., Jadud, M. C., Amarra, A. C. M., Dy, T., Espejo-Lanoz, M. B. V., ..., Tabanao, E. S. (2009). Affective and behavioral predictors of novice programmer achievement. In *Proceedings of the 14th Annual ACM SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE '09)*, 6–9 July 2009, Paris, France (pp. 156–160). New York, NY: ACM. <https://dx.doi.org/10.1145/1562877.1562929>
- Rodrigo, M. M. T., Baker, R., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., ..., Viehland, N. J. B. (2007). Affect and usage choices in simulation problem solving environments. In *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007)*, 9–13 July 2007, Los Angeles, CA, USA (pp. 145–152). The Netherlands: IOS Press. Retrieved from <https://dl.acm.org/doi/10.5555/1563601.1563629>
- Rodriguez, F., Yu, R., Park, J., Rivas, M. J., Warschauer, M., & Sato, B. K. (2019). Utilizing learning analytics to map students' self-reported study strategies to click behaviors in STEM courses. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK '19)*, 4–8 March 2018, Tempe, AZ, USA (pp. 456–460). New York, NY: ACM. <https://dx.doi.org/10.1145/3303772.3303841>
- Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research*, 21(4), 37–59. [https://dx.doi.org/10.1300/J079v21n04\\_02](https://dx.doi.org/10.1300/J079v21n04_02)
- Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York, NY: McGraw-Hill.
- Rozin, P., & Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, 3, 68–75. <https://dx.doi.org/10.1037/1528-3542.3.1.68>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://dx.doi.org/10.3102/0034654307313795>
- Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 3(1), 48–51. <https://dx.doi.org/10.1037/a0014632>
- 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. <https://dx.doi.org/10.1037/a0017081>
- Sreerekha, S., Arun, R. R., & Sankar, S. (2016). Effect of predict-observe-explain strategy on achievement in chemistry of secondary school students. *International Journal of Education & Teaching Analytics*, 1(1). Retrieved from <http://www.innoriginal.com/index.php/ijeta/article/view/35>
- Storm, C., & Storm, T. (1987). A taxonomic study of the vocabulary of emotions. *Journal of Personality and Social Psychology*, 53, 805–816. <https://dx.doi.org/10.1037/0022-3514.53.4.805>
- Sun, P. (2006). *Outlier detection in high dimensional, spatial and sequential data sets* (PhD thesis). Sydney University.
- Tao, P. K., & Gunstone, R. F. (1999). The process of conceptual change in force and motion during computer-supported physics instruction. *Journal of Research in Science Teaching*, 36(7), 859–882. [https://dx.doi.org/10.1002/\(SICI\)1098-2736\(199909\)36:7<859::AID-TEA7>3.0.CO;2-J](https://dx.doi.org/10.1002/(SICI)1098-2736(199909)36:7<859::AID-TEA7>3.0.CO;2-J)
- Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, 8(6), 951–970. <https://dx.doi.org/10.1007/s10763-010-9203-x>
- VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209–249. [https://dx.doi.org/10.1207/S1532690XCI2103\\_01](https://dx.doi.org/10.1207/S1532690XCI2103_01)
- White, P. A. (1989). Evidence for the use of information about internal events to improve the accuracy of causal reports. *British Journal of Psychology*, 80, 375–382. <https://dx.doi.org/10.1111/j.2044-8295.1989.tb02327.x>
- White, R. F., & Gunstone, R. T. (1992). *Probing understanding*. London, UK: Routledge. <https://dx.doi.org/10.4324/9780203761342>
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27(4), 551–572. [https://dx.doi.org/10.1016/S0361-476X\(02\)00006-1](https://dx.doi.org/10.1016/S0361-476X(02)00006-1)
- Wise, A., Speer, J., Marbouti, F., & Hsiao, Y. T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, 41, 323–343. <https://dx.doi.org/10.1007/s11251-012-9230-9>
- Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, 22(6), 413–419. <https://dx.doi.org/10.1016/j.learninstruc.2012.03.004>