

Modeling confusion in collaborative learning

Peter Bagdovitz

Western Washington University, Bellingham, WA, 98225, USA

Jayson M. Nissen

Nissen Education Research and Design, Slidell, LA, 70458, USA

When students experience confusion, resolving that confusion can lead to deeper understanding and engagement. Persistent unresolved confusion, however, can lead students to frustration and disengagement. Our research explores confusion and other emotions associated with learning as students work through elicit confront resolve (ECR) activities in Next Generation Physical Science and Everyday Thinking (Next Gen PET) physics courses for pre-service elementary teachers. We used the experience sampling method (ESM) to measure students' subjective experiences during seven particularly confusing activities. The ESM asked about confusion, self-efficacy, engagement, and stress, which we chose to align with existing models of confusion in learning. After some revision, our model fit the data well using confirmatory factor analysis. The only activity that required students' consistent use of mathematics produced the highest levels of confusion and stress for students. This relationship with mathematics content indicates the Next Gen PET courses could better support pre-service elementary teachers developing fluency and comfort with mathematics.

I. INTRODUCTION

While students and instructors often perceive confusion as a stumbling block in the learning process, confusion acts as a catalyst for growth and development [1]. As students navigate through the complexities of physics courses, they must confront confusion often. When students resolve their confusion, they both learn and gain a sense of achievement, but being stuck and confused can lead to stress and disengagement. To support students learning through complex tasks, research-based materials often have students work in pairs or small groups. Student groups' ability to manage confusion and the attendant emotions that may accompany it varies. Some groups maintain a productive focus on collaborative physics learning, while confusion may derail other groups and lead to frustrated and disengaged students.

Physics educators often use collaborative learning strategies [2, 3] in the design and implementation of research-based, student-centered instruction. Student-centered instruction in physics, such as Tutorials in Introductory Physics [4] and Next Generation Physical Science and Everyday Thinking (Next Gen PET) [5], often uses an elicit confront resolve (ECR) framework. In an ECR framework, classroom activities present students with scenarios known to produce confusion and students work together to resolve their confusion. Curricular materials typically scaffold this resolution; however, they focus on conceptual understanding with little explicit support for students in managing social interactions or emotions.

To investigate these confusing experiences, we used the D'Mello *et al.* [1] model of affect shown in Fig 1. We adopt D'Mello's description of confusion as an epistemic or a knowledge emotion. Confusion arises due to information-based assessments of how incoming information aligns with existing knowledge structures, as well as inconsistencies within the incoming information [1]. This model led us to measure stress, self-efficacy, engagement, and confusion during these activities. D'Mello and colleagues' model applies to students engaged in learning activities who reach an impasse, such as in ECR-based activities. When students detect an impasse they may become confused. If they become confused they can then resolve their confusion through problem-solving and teamwork and continue to engage in the activity. If the activity was very confusing, resolving the confusion may create a sense of accomplishment and self-efficacy. Failure to resolve confusion can lead to stress and eventually disengagement. Investigating these emotions can identify topics or activities that elicit confusion and inform the extent to which the curricular materials support students in learning from their confusion or fail to support the students.

II. RESEARCH QUESTIONS

Our research was driven by two questions.

1. How well does the proposed factor structure of the

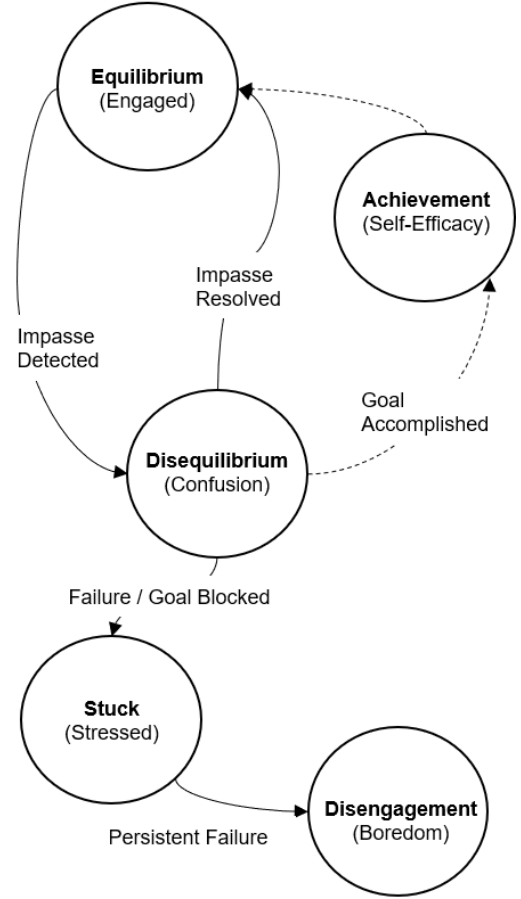


FIG. 1: [1] Model of productive and unproductive confusion.

items on the Experience Sampling Method (ESM) questionnaire align with the collected data?

2. What activities and other emotions are associated with higher levels of confusion?

The relationships between activities, confusion, and other emotions can allow us to identify activities where students failure to resolve their confusion leads to stress and disengagement. Identifying these confusing activities will assist our broader project on confusion and inform changes to the Next Gen PET curriculum to improve student cognitive and affective outcomes and collaboration.

III. METHODS

The research occurred in introductory physics courses for pre-service elementary educators. The courses used the Next Gen PET curriculum [5] and were taught at two primarily undergraduate institutions on the west coast of the United States. We collected data during 2021 and 2022 in nine courses taught by six different instructors. The summer 2021 courses were taught online and all other courses were taught in person.

The Next Gen PET curriculum prepares students to teach elementary-level science by covering a wide range of physics topics and focusing on students developing scientist-like views of these physics topics [6]. In the classes where we conducted the research, students worked in groups of 3-4. The students worked through activities that were part of the units for energy-based models (UEM), force-based models (UFM), and combination of forces (UCF). These units are divided into activities designed to cover one class period (e.g., UEM-A6 or UCF-A2).

We used the experience sampling method (ESM) [7–9] to measure students’ subjective experiences on four latent variables and one question on confusion. One-hundred and sixty students completed 650 ESM surveys. As a part of the ESM, the students completed brief surveys during activities that the course instructors and members of the research team identified as especially confusing for students. The research team marked the class activities with stickers to prompt the students to complete the surveys. At one institution, the stickers included a QR code that directed students to complete the survey online. At the other institution, the stickers prompted them to complete and turn in a paper survey that was inserted into the activity. The research used paper surveys because some students may not have had a device to access an online survey in class.

The ESM data collection was part of a larger systematic investigation of students’ confusion and socio-metacognition [10] in Next Gen PET courses. All students completed the ESM during class and a reflection prompt about the course outside of class. Researchers video recorded two to three groups of students in each section. Students from these recorded groups participated in interviews on their experiences. Other researchers in the project are analyzing these other data streams, but that data will not be discussed in this paper.

We compared four emotions across seven activities the instructors identified as especially confusing for students. These four emotions were stress, self-efficacy, confusion, and engagement. We also collected data on teamwork, but we did not include that measurement because student responses all had ceiling effects with little variation. To compare the activities, we used descriptive statistics, data visualizations, and the Wilcoxon rank sum test. These methods allowed us to identify the extent to which students experienced confusion in each activity and to identify an activity that elicited more confusion than the other activities. We then compared how stress, engagement, and self-efficacy compared in the activity where students reported high levels of confusion to the other activities. The Wilcoxon rank sum test is a non-parametric alternative to the paired t-test for independent samples. Researchers can use an equivalent to correlation, r , as an effect size [11]. The data visualizations overlaid a plot of the individual data points on top of the box plots.

A. CFA

We conducted confirmatory factor analysis (CFA) [12] to test the construct validity of our ESM survey. CFA is a type of structural equation modeling that tests the correlations in a model proposed by the researcher. We used a three-step CFA process to come to a model that fit the data well. In the three steps we (i) created an initial model based on theory and prior research, (ii) ran CFA and generated fit statistics using the lavaan [13] package, and (iii) used the modindices command to improve the fit. We repeated these last two steps until the factor loadings in the model were > 0.6 and the fit indices passed the cutoffs discussed in the next paragraph. Hair *et al.* [14] proposes factor loadings of 0.5, explaining 25% of the variance in the item, as an absolute minimum and 0.7 as a preferred minimum, 50% explained variance.

We used the root-mean-square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI) to inform how well the factor structure fits the data. RMSEA is an absolute fit index that addresses parsimony in the model by accounting for the degrees of freedom in the model. High RMSEA indicates an over-constrained model with too few degrees of freedom. RMSEA has a cutoff of 0.06 and below [15]. CFI and TLI are relative fit indices. They compare the test model to a baseline model with all covariances set to zero and all variances freely estimated. In other words, CFI and TLI compare the test model to a very poor model, so they look very good for models that fit the data reasonably well. Both indices range from 0 to 1, where 1 is the best possible fit and 0 indicates no fit. We used a cutoff of CFI and TLI > 0.95 [15]. A combination of absolute (RMSEA) and relative (CFI and TLI) fit indices allowed us to increase the fit of the model without over complicating the model. This balance leads to a model that is simple and fits well.

IV. RESULTS

Our research was driven by two questions. How well does the proposed factor structure of the items on the ESM align with the collected data, and what activities and other emotions are associated with higher levels of confusion? To answer the first question, we discuss the CFA and path diagram. We then use the latent factors for confusion, stress, self-efficacy, and engagement to investigate students’ confusion.

The iterative process for improving our CFA model produced a factor structure (Fig. 2) that fit the ESM data well. The standardized factor loadings for each item exceeded our cutoff of > 0.6 . Fit indices also passed our cutoffs of RMSEA < 0.06 , CFI and TLI > 0.95 .

The initial model differed from the model shown in 1 in that the initial model included a question on how in control students feel under the latent factor for *self-efficacy* and a question on how much students wanted to do the activity under the latent factor for *engagement*. The factor loadings

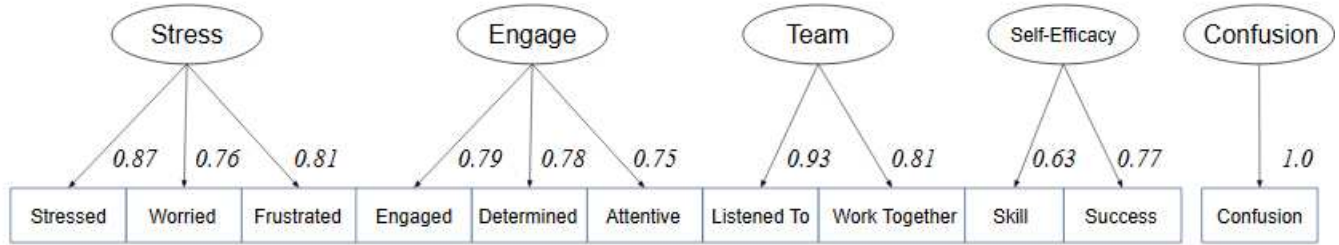


FIG. 2: Diagram of the final factor structure. The boxes are questions on the ESM, the ovals are latent factors, and the factor loadings are the arrows linking them.

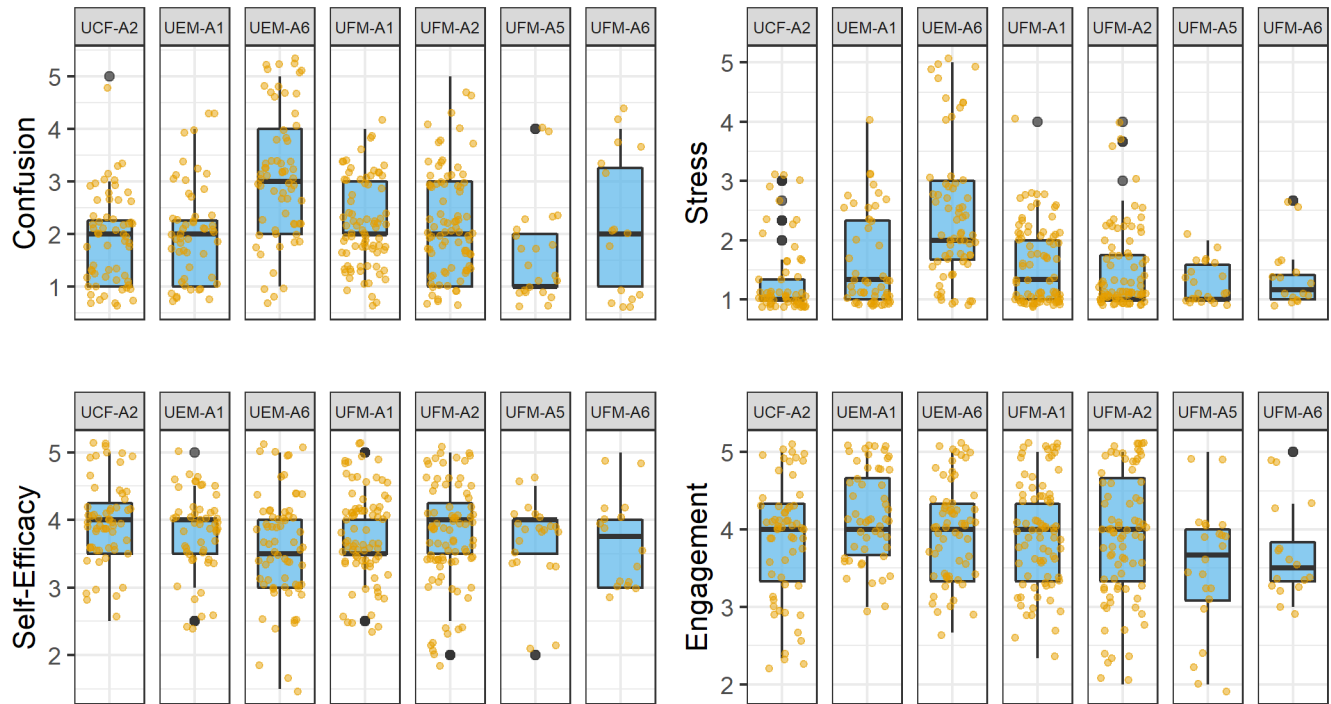


FIG. 3: Confusion, stress, self-efficacy and engagement across all seven activities. The scale ranged from 1, not at all, to 5, extremely. The lighter points in gold represent the individual responses from students with a slight 'jitter' to space them apart. The solid black dots are outlier responses for the boxplot. The activities are curricular materials meant to be covered in one course period and following the naming convention in the curriculum.

for these two items were below the cutoff of 0.6. We first removed the *want to do* question from the model and then removed the *control* question.

The following results revealed the activities and emotions that are associated with higher levels of confusion. Figure 3 presents students' reported confusion, stress, self-efficacy, and engagement in seven activities. In most activities, most students experienced low levels of confusion. Median scores for confusion were between 1, not at all, and 2, somewhat, for all activities besides the sixth unit of the energy section of the class, UEM-A6.

The confusion in UEM-A6 stands out from the other activities. In UEM-A6 students experienced a median confusion of 3; the typical student participating in this activity was mod-

erately confused. The interquartile range for UEM-A6 does not extend to 1, so few students worked through the activity without becoming confused at all. A large cluster of students, shown in gold in Fig. 3, responded as experiencing extreme confusion in UEM-A6, more than in all other activities combined. A Wilcoxon signed rank test of confusion in UEM-A6 versus confusion in all other activities showed that this difference was statistically significant, $p < 0.001$, with an effect size of $r = 0.27$, with confidence intervals of $[0.19, 0.35]$.

Comparing the stress, self-efficacy, and engagement in UEM-A6 to the other activities reveals some patterns consistent with the D'mello model and some inconsistencies. Stress was higher in UEM-A6 with a median of 2 compared to 1 or 1.3 in other activities, and more students reported very or ex-

treme stress in UEM-A6 than in all other activities; $p < 0.01$, $r = 0.31$ [0.23, 0.38]. Self-efficacy was lower than in most other activities. This difference was not as large as for confusion or stress. The difference in self-efficacy compared to all other activities was small $p = 0.06$, $r = 0.07$ [0.00, 0.15]. Students reported similar engagement in UEM-A6 to the other activities, $p = 0.33$, $r = 0.04$ [0.00, 0.11], and they did not report the disengagement the D’Mello model indicates high levels of confusion can lead to, see Fig. 1.

V. DISCUSSION

We studied seven activities instructors identified as particularly confusing, but as Fig. 3 shows students tended to report ‘slightly’ to ‘not at all’ confused in all activities besides UEM-A6. This contrast between students reporting little to no confusion and instructors identifying these activities as particularly confusing may indicate a low level of metacognition for the students, or it could indicate that instructors identified activities where some students were very or extremely confused. This difference between instructors and students could, however, have resulted from a difference between what the Next Gen PET materials expect students to do and what instructors with PhDs in physics think of as a full understanding of these topics. Further exploring students’ metacognition in these activities can inform the extent to which students engage in metacognition and the ways instructors and curricula can scaffold students’ metacognition.

Descriptive statistics indicated the ECR activities aligned with many aspects of the D’Mello *et al.* [1] model of affect. In the UEM-A6 activity, where students reported the highest levels of confusion, students also reported higher stress but only slightly lower self-efficacy. The elevated stress, especially in activity UEM-A6, indicates some students did not resolve the confusion elicited by the activity and likely failed to achieve a sense of accomplishment.

In UEM-A6, students did calculations involving fractions and percentages in each step of the activity. Students could not circumvent the mathematical sections. Only UEM-A6 had significant mathemats required in the activity.

We cannot conclude how engagement correlates with confusion within this study. Because the ESM only captured a snapshot of students’ experiences, the data cannot clarify if they stayed engaged over the whole activity or disengaged at a later point. To address this limitation of the ESM, we are also using classroom videos, interviews, and journal reflections to study the confusion, stress, and engagement students experienced in Next-Gen PET with a focus on UEM-A6 as the most confusing activity. For example, one student who found UEM-A6 very confusing and stressful further engaged

with the material by seeking help from instructors outside of class.

VI. CONCLUSION

The Next Gen PET curriculum seldom requires students to do mathematics. UEM-A6, however, required students to do mathematics throughout the activity. The higher levels of stress and confusion that students experienced in UEM-A6 than the other confusing activities indicates that the mathematics content and students’ mathematics anxiety may have been the primary source of confusion and stress in this activity. Next Gen PET courses primarily serve pre-service elementary school teachers. Elementary school teachers tend to have higher levels of mathematics anxiety than any other professions [16], and they experienced their highest levels of mathematics anxiety in university [17]. Next Gen PET instructors could integrate more mathematics into their activities to support students in developing confidence and fluency for mathematics. Supporting pre-service teachers in developing this fluency and confidence would likely increase the mathematics learning of their future students [18].

The D’Mello model [1] focuses on several emotions students experience while learning and experiencing confusion, but none of these affective emotions address students’ pre-dispositions towards a course or content area, such as mathematics anxiety. As shown in Fig. 1, the D’Mello model posits the student begins in an equilibrium state characterized by a feeling of engagement, but a student experiencing mathematics anxiety may begin an activity with similar engagement to their peers but with higher stress levels than their peers. They may even feel stuck or stressed once they realize the activity requires them to do mathematics. The D’Mello model implies that students experiencing this elevated stress may be more likely to perceive themselves as stuck and shift to disengagement than they would in other activities or than their less anxious peers. Students with higher levels of mathematics anxiety likely need more support and scaffolding from their peers and instructors and the curriculum to resolve their confusion, gain a sense of achievement and self-efficacy, and maintain their engagement in the activity.

VII. ACKNOWLEDGEMENTS

This work was made possible by funding from the National Science Foundation (Grant No. 1928596). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also thank Thanh Lê, Andrew Boudreaux, and Carolina Alvarado for their work on the project.

[1] S. D’Mello, B. Lehman, R. Pekrun, and A. Graesser, Confusion can be beneficial for learning, *Learning and Instruction* **29**, 153

(2014).

- [2] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *American Journal of Physics* **66**, 64 (1998).
- [3] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proceedings of the National Academy of Sciences* **111**, 8410 (2014).
- [4] L. C. McDermott, P. S. Shaffer, *et al.*, *Tutorials in introductory physics* (Prentice Hall Upper Saddle River, NJ, 2002).
- [5] F. Goldberg, S. Robinson, E. Price, D. Harlow, J. Andrew, and M. McKean, *Next generation physical science and everyday thinking*, Greenwich, CT: Activate Learning (2018).
- [6] V. K. Otero and K. E. Gray, Attitudinal gains across multiple universities using the physics and everyday thinking curriculum, *Physical Review Special Topics-Physics Education Research* **4**, 020104 (2008).
- [7] J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi, *Experience sampling method: Measuring the quality of everyday life* (Sage, 2007).
- [8] J. M. Nissen and J. T. Shemwell, Gender, experience, and self-efficacy in introductory physics, *Physical Review Physics Education Research* **12**, 020105 (2016).
- [9] J. M. Nissen, Gender differences in self-efficacy states in high school physics, *Physical Review Physics Education Research* **15**, 013102 (2019).
- [10] M. Borge and B. White, Toward the development of socio-metacognitive expertise: An approach to developing collaborative competence, *Cognition and Instruction* **34**, 323 (2016), <https://doi.org/10.1080/07370008.2016.1215722>.
- [11] M. Tomczak and E. Tomczak, The need to report effect size estimates revisited. an overview of some recommended measures of effect size., *Trends in Sport Sciences* **21** (2014).
- [12] K. A. Bollen, *Structural equations with latent variables*, Vol. 210 (John Wiley & Sons, 1989).
- [13] Y. Rosseel, lavaan: An R package for structural equation modeling, *Journal of Statistical Software* **48**, 1 (2012).
- [14] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. Tatham, *Data Analysis Multivariate* (New Jersey. Upper Saddle River. Pearson Education, 2006).
- [15] P. Eaton and S. D. Willoughby, Confirmatory factor analysis applied to the Force Concept Inventory, *Physical Review Physics Education Research* **14**, 010124 (2018).
- [16] R. Hembree, The nature, effects, and relief of mathematics anxiety, *Journal for Research in Mathematics Education* **21**, 33 (1990).
- [17] D. Patkin and Y. Greenstein, Mathematics anxiety and mathematics teaching anxiety of in-service and pre-service primary school teachers, *Teacher Development* **24**, 502 (2020).
- [18] G. Ramirez, S. Y. Hooper, N. B. Kersting, R. Ferguson, and D. Yeager, Teacher math anxiety relates to adolescent students' math achievement, *AERA open* **4**, 2332858418756052 (2018).