Citation: Sullins, J., Console, K., Denton, R., Henrichson, C., & Barber, S. (in press). Not all confusion is productive: An investigation into confusion induction methods and their impact on learning. *International Journal of Learning Technology.*

Not all confusion is productive: An investigation into confusion induction methods and their impact on learning

Jeremiah Sullins, Katie Console, Rebecca Denton, Clayton Henrichson, Steven Barber

Harding University, U.S.A.

Author Note

Jeremiah Sullins, Department of Behavioral Science, Harding University

Katie Console, Department of Behavioral Science, Harding University

Rebecca Denton, Department of Behavioral Science, Harding University

Clayton Henrichson, Department of Behavioral Science, Harding University

Steven Barber, Department of Behavioral Science, Harding University

Correspondence concerning this article should be addressed to Dr. Jeremiah Sullins, Department of Behavioral Science, Harding University, Searcy, AR, 72149

Contact: jsullins@harding.edu

Abstract

The current study was an attempt to discover the gold standard of inducing a state of confusion that is beneficial to the learning of complex science topics. Using a randomized controlled trial, participants received either one of three different types of confusion induction (deep-questions, intra-testing, breakdown scenarios) or a lecture based information delivery control. Results revealed that breakdown scenarios were the most beneficial in terms of pretest to posttest learning gains. Additionally, significant interactions were discovered among learning, confusion induction methods, and measures of individual differences (i.e., goal orientation and attributional complexity). Interpretations and applications are discussed.

Keywords: affective sciences, productive confusion, individual differences, STEM learning

Not all confusion is productive: An investigation into confusion induction methods and their impact on learning

According to the National Assessment of Educational Progress in the United States, 4th grade students scored a 154 (out of 300) in science in 2015 which is up from 150 (out of 300) back in 2011. Additionally, 8th grade scored a 154 (out of 300) in science compared to 152 (out of 300) back in 2011 (NAEP, 2015). Although these results are encouraging and clearly heading in the right direction, the news is not all good. According to the NAEP, in 2015 there was no change from 2011 in 12th grade students' science scores. Furthermore, in 2015, 78% of twelfth graders scored below "proficient" in science. These results are suggesting that teachers still need more effective methods of instructing students in the STEM (science, technology, engineering, and math) disciplines. This raises the question: Are there effective methods of teaching STEM content that can be proven empirically, which can also be easily transferred into the classroom? Although it is beyond the scope of this paper to explore all possible methods of STEM instruction, there is empirical evidence (in and out of the laboratory) suggesting that emotions are inextricably linked to learning. More specifically, recent research has suggested that "productive confusion" can significantly increase learning gains when compared to various controls. To clarify, productive confusion is an instance of confusion that can be immediately or eventually resolved. Compare this to an instance of "hopeless confusion" in which learners are unable to resolve their experience of confusion. This paper seeks to answer the following questions:

RQ1: *What are the most efficient methods of inducing productive confusion in order to increase STEM learning gains?*

RQ2: *What role and to what extent do individual differences play in regards to confusion induction and learning of complex science topics?*

Confusion and Learning

Research has not only discovered that emotions are present during learning, but certain emotions are inextricably linked to learning. In fact, the 21st century has been ripe with empirical evidence exploring links between emotions and cognition (D'Mello & Millis, 2014; D'Mello, Lehman, Pekrun, & Graesser, 2014; Lehman, D'Mello, Strain, Millis, Gross, Dobbins, Wallace, Millis, & Graesser, 2013). For example, Pekrun (2010) refers to the term "academic emotions" when referring to the emotions that learners experience during a broad range of educational activities. Other researchers have explored what are called "learning-centered emotions" (Calvo & D'Mello, 2011; Rodrigo & Baker, 2011). These learning-centered emotions consist of anxiety, boredom, confusion, curiosity, engagement/flow, frustration, happiness, and surprise. The focus of this paper is on one specific learning-centered emotion: confusion. The reason that we are focusing on confusion is because research has suggested that confusion is both prevalent in and important to learning.

Imagine a scenario in which a student signs up for an online physics course. As this student progresses through the material at his/her own pace, he/she encounters videos and illustrations. Typically, when a student comes across a situation such as this, he/she may pay attention to the information being delivered. This alone is not always sufficient for learning to occur. Learners may watch the instructional video and report that they have understood the content presented in the video. However, learners may not have sufficiently engaged with the content of the video, which can lead to simply reinforcing the previously held misconceptions of the learner. This issue can be resolved fairly easily by simply embedding videos that directly address the prominent misconceptions that students hold about a certain topic (Muller, 2012). This fairly simple intervention creates greater uncertainty for learners because they must reconcile their previously held misconceptions and the information being presented, an effortful process which can result in deeper learning.

Craig, Graesser, Sullins, and Gholson (2004) conducted a study in which participants interacted with an intelligent tutoring system on the topic of computer literacy. Results revealed that three specific emotions were related to learning. Boredom and frustration were found to be negatively correlated with learning. However, confusion was positively correlated with learning. Follow up tests revealed that only confusion was predictive of learning with more instances of confusion significantly predicting increased learning from students. This finding may initially seem counterintuitive due to the assumption that negatively-valenced emotions would negatively impact learning, as was the case for boredom and frustration. Subsequent research has replicated the finding that confusion occurs frequently during and can actually be beneficial to learning (Baker et al., 2010; Craig et al., 2004; D'Mello & Graesser, 2011; Graesser, Chipman et al., 2007; Rodrigo & Baker, 2011; Lehman et al., 2008; D'Mello, Lehman, & Person, 2010; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003).

The relationship between confusion and learning is also consistent with several theories that highlight the merit of impasses during learning (Brown & VanLehn, 1980; VanLehn et al. 2003) and with theories that claim that cognitive disequilibrium is one precursor to deep learning (Graesser & Olde, 2003; Piaget, 1952). According to these theories, confusion is triggered when learners are confronted with information that is inconsistent with existing knowledge and learners are unsure about how to proceed. These events that trigger impasses place learners in a state of cognitive disequilibrium, which is ostensibly associated with heightened physiological arousal and more intense thought as learners attempt to resolve impasses.

Ways of Inducing Productive Confusion

If confusion under certain circumstances is desirable, then what is the best way to induce this desirable confusion? Although the results are equivocal thus far, there has been research exploring different methods of producing a state of productive confusion in learners. Educational research on cognitive conflict can shed light on potential methods to induce productive confusion (Limon, 2001). Cognitive conflict, a state that is similar to confusion and cognitive disequilibrium, is triggered when an individual is confronted with discrepant events, such as deviations from the norms, obstacles to goals, interruptions of action sequences, contradictions, anomalous information, unexpected feedback, and other forms of uncertainty. Limon (1995; as cited in Limon 2001), presented anomalous information to two groups of individuals who had high prior knowledge of historical problems. Results revealed that the presentation of anomalous data allowed the learners to develop a more elaborate and sophisticated response to questions. Furthermore, Limon and Carretero (1997; as cited in Limon 2001), also showed that the presentation of anomalous information may be beneficial not only for learners with high prior knowledge but also for those that have low domain knowledge of a particular topic. Limon proposes that although anomalous information may not lead to radical change in the students' thinking, it may serve as a catalyst in the process of conceptual change.

Furthermore, Graesser and McMahen (1993) explored question asking while participants read expository texts with varying degrees of anomalous information. The results from this study suggest that having students read text with *anomalous information* that presented an obstacle may have placed the students in cognitive disequilibrium, which in turn significantly increased higher quality questions.

More recently, Lehman, D'Mello, and Graesser (2012) explored the complex relationship that exists between confusion and learning gains. Averaged across five different studies, the emotions that were most prevalent were: confusion (17%), frustration (13%), boredom (18%), engagement/flow (24%), delight (6%), surprise (3%), and neutral (19%). As can be seen from these results, confusion is a prevalent emotion during learning complex science. Further data mining revealed that certain elements of the tutorial dialogue were related to the occurrence of confusion (D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; D'Mello, Craig, Sullins, & Graesser, 2006). Confusion was found to occur when the tutor is less direct (hinting and prompting for information) and when the tutor provides negative feedback. These results suggest that the implementation of particular pedagogical strategies (e.g., difficult problems, hints, prompts) can induce confusion, which, if productively managed, can result in improved learning. Additional research has explored the use of contradictory information during the learning of critical thinking skills (Lehman et al., 2013; D'Mello et al., 2014). Specifically, the researchers tested the prediction that contradictory information can be a useful way of inducing confusion within the context of information presented via animated pedagogical agents. The results revealed that contradictions did successfully evoke confusion. However, the magnitude of confusion was dependent upon the source and severity of the contradiction.

The results of the aforementioned studies demonstrate that there are a multitude of instructional methods that at some level produce a state of confusion in the learner. More specifically, methods such as leaving out critical information, presenting contradictions or adding puzzling information, vague hints and prompts, or the presentation of anomalous information.

However, simply placing a learner in a state of confusion may not be sufficient enough to promote deep conceptual change within the learner. For example, Limon (2001) in a review of the cognitive conflict literature expresses concern over the fact that in the majority of cases, producing a state of cognitive conflict in a learner does not lead to a strong restructuring and consequently a deep understanding of the new information. This lack of effectiveness could be due to cases of confusion being left unresolved therefore leading to a path of negative affect (e.g., frustration and boredom) which in turn could lead to negligible learning gains.

Resolution of Confusion

Inducing a state of confusion in a learner only produces superficial learning gains in a large amount of cases. This superficial learning can be compared to Piaget's beta level of adaptive responses in which learners only show partial modification of their understanding of a particular topic. One possible explanation for this lack of deep conceptual change could be due to the fact that the confusion that learners experience during these various learning tasks is never fully resolved. This unresolved confusion could lead to a state of hopeless confusion (as opposed to productive confusion), which results in an unsuccessful cycle of frustration and boredom. According to D'Mello and Graesser (2012), frustration occurs when an individual experiences repeated failures and is stuck. Persistent (or hopeless) confusion occurs when conflict resolution fails and an individual is unable to restore equilibrium. Conflict resolution requires people to stop, think, effortfully deliberate, problem solve, and revise their existing mental models (see Figure 1 for a theoretical explanation of productive versus hopeless confusion).

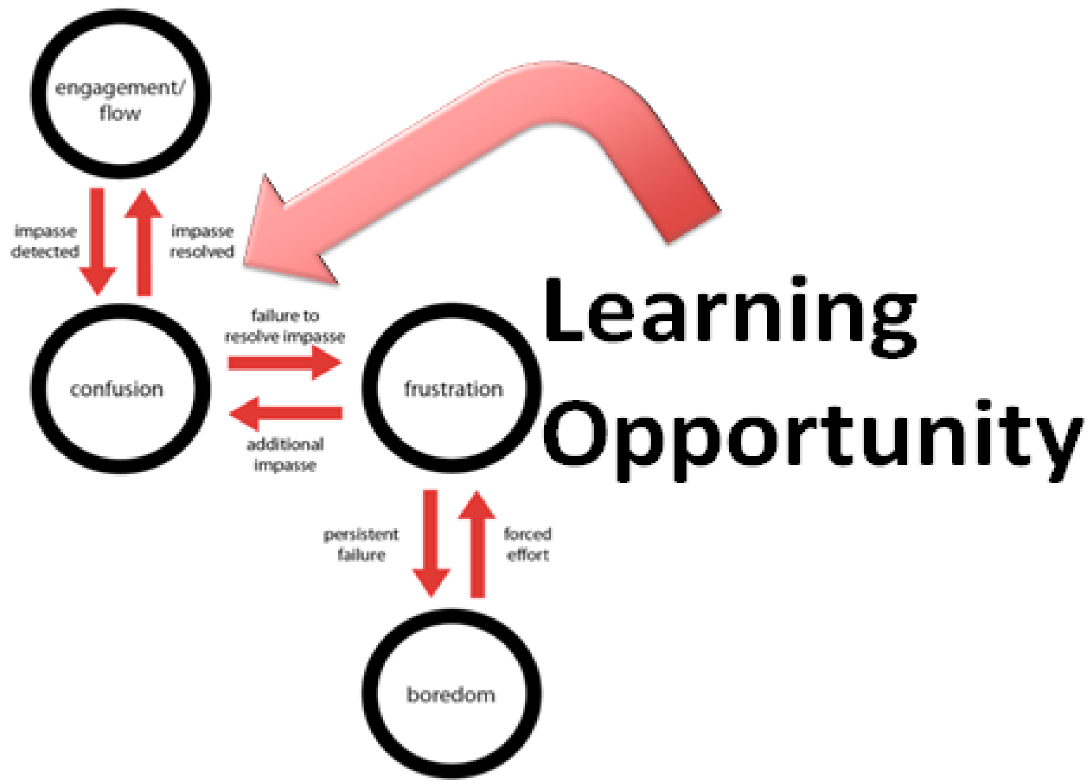

Figure 1. Vicious versus virtuous learning cycle

There is some evidence (albeit limited) stating the importance of conflict resolution (Lee, Rodrigo, Baker, Sugay, & Coronel, 2011). D'Mello and Graesser (2012) explored conflict resolution within the context of device breakdowns. Participants were presented with an illustrated text of a household device (e.g., cylinder lock) along with a description of the device breakdown (i.e., inducing confusion). In order to resolve confusion, a learner must search for the causes of the breakdown and find ways to fix the device. Successful conflict resolution will result in an extended mental model that accommodates the device breakdown. Results revealed that the participants who were in the device breakdown condition who also at least in part resolved their confusion significantly outperformed their unresolved counterparts on tests of device comprehension and processing of the breakdown scenarios.

These studies suggest that hopeless confusion can lead to a student getting "trapped" in the negative affect cycle of frustration and boredom. More specifically, if students are unable to work through their confusion, they are likely to give incorrect responses to questions which eventually can lead to the student giving up or attempting to "game the system" when possible.

Individual Differences

Individual differences have been found to play a substantial role in the domain of learning and emotions (Sullins & Goza, 2013; Sullins, Goza, Smith, Moore, & Morrow, 2014; Sullins & Graesser, 2014). Limon (2001) states that researchers have focused exclusively on confusion inducing materials and largely have neglected the importance of learner individual differences. The impact of individual difference is crucial due to the fact that there is clearly no one-size-fits-all approach to learning. By investigating confusion induction through device breakdowns, D'Mello and Graesser (2014) found that successful confusion resolution varied based on learners' scholastic aptitude.

Furthermore, Sullins and Graesser (2014) explored student question generation as a result of being placed in a state of cognitive disequilibrium. The analyses revealed that motivation and personality traits (e.g., agreeableness and neuroticism) were significant predictors of instances of confusion.

Finally, Lehman, D'Mello, and Graesser (2013) through a cluster analysis discovered that students differed with respect to cognitive ability and cognitive drive. Students with a combination of high cognitive ability and high cognitive drive benefited the most from the confusion induction (i.e., false feedback) and performed better on tests of transfer knowledge. Although it is beyond the scope of this paper to address all instances in which individual differences have been shown to be germane to learning, emotions, and confusion, it is clear that this is an area that warrants further investigation.

In summary, the apparent benefits of emotions, inducing confusion, the resolution of confusion, and individual differences involved in confusion beg the following research questions:

RQ1: *What are the most efficient methods of inducing productive confusion in order to increase STEM learning gains?*

RQ2: *What role and to what extent do individual differences play in regards to confusion induction and learning of complex science topics?*

Current Study

Participants

Participants consisted of 178 students enrolled in an introduction to psychology course at a private liberal arts university located in the southern United States. Participants were given extra credit for their participation.

Methods and Procedure

In this study, we conducted a laboratory-based randomized experimental trial with **four levels** of our independent variable (i.e., induction of desirable confusion): **1) Breakdown Scenarios 2) Deep Questions 3) Intra-testing, and 4) Control** (i.e., information delivery) to determine impact on students' learning.

After entering the lab, participants completed three tests of individual differences: Achievement Goal Questionnaire, Attributional Complexity Scale and Dweck's Mindset Scale.

The first measure of individual differences administered was the Achievement Goal Questionnaire (Elliot & McGregor, 2001). The Achievement Goal Questionnaire measures the following attitudes: (a) Performance Approach (alpha = .92) in which a student seeks to demonstrate or prove competence, particularly in the presence of an audience (e.g., "my goal in this class is to get a better grade than most of the other students"), (b) Performance Avoidance (alpha = .83) in which a student seeks to demonstrate or prove that he or she is not incompetent, particularly in the presence of an audience (e.g., "I just want to avoid doing poorly in this class"), (c) Mastery Avoidance (alpha = .89) in which the student seeks to better understand the material regardless of an external outcome (e.g., grades) but does so to avoid appearing less competent in front of others (e.g., sometimes I am afraid that I may not understand the content of this class as thoroughly as I'd like"), and (d) Mastery Approach (alpha = .87) in which the student seeks to better understand the material regardless of an external outcome in order to appear competent (e.g., I desire to completely master the material presented in this class).

Participants then completed the Attributional Complexity Scale which was designed to investigate the different ways that people think about themselves and other people. The scale measures seven attributional constructs: 1) a motivational component 2) preference for complex rather than simple explanations 3) metacognition concerning explanations 4) awareness of the extent to which people's behavior is a function of interaction with others 5) a tendency to infer abstract or causally complex internal attributions 6) a tendency to infer abstract, contemporary, and external causal attributions, and 7) a tendency to infer external causes operating from the past (Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986).

Following the Attributional Complexity Scale, participants were administered the Dweck's Mindset Scale. This scale explored any influences of learners' mindsets on desirable confusion. More specifically, Dweck proposed that learners have either a fixed mindset or growth mindset. In a fixed mindset, learners believe that their intelligence is a fixed quantity.

Learners with a fixed mindset tend to agree with survey items like "you have a certain amount of intelligence, and there is not much you can do to change that." Other learners believe that they have a growth mindset. In a growth mindset learners believe that intelligence has the potential to grow and improve in response to effort, good strategies, and help from others. In terms of confusion and cognitive disequilibrium, this can translate to whether students engage with difficult, confusing material and work to resolve their confusion or simply give up on the task and move on to something else. For example, a student with a growth mindset would be more likely to view an instance of confusion as a challenge or jumping off point to learn more and increase their understanding of the topic and subsequently their intelligence.

In order to assess learning, a physics pretest was administered following the tests of individual differences. Two different versions of a multiple choice test were used (pretest and posttest; 13 questions each). Because productive confusion is designed to strengthen the understanding of how things work and why things are the way they are (opposed to the memorization of facts), these multiple choice tests were designed to test deep-level knowledge. The tests consisted of physics information that will be covered during the learning sessions. These physics tests were developed by the researchers associated with the proposed project (i.e., a physics professor at the university where the study was conducted).

Participants were then randomly assigned to one of **four levels** of our independent variable (i.e., induction of desirable confusion): **1) Breakdown Scenarios 2) Deep Questions 3) Intra-testing, and 4) Control** (no confusion induction method) to determine impact on students' learning. In all sessions, the participants watched pre-recorded "tutoring/content delivery" (See Figure 2). Physics was the content domain used in all conditions.

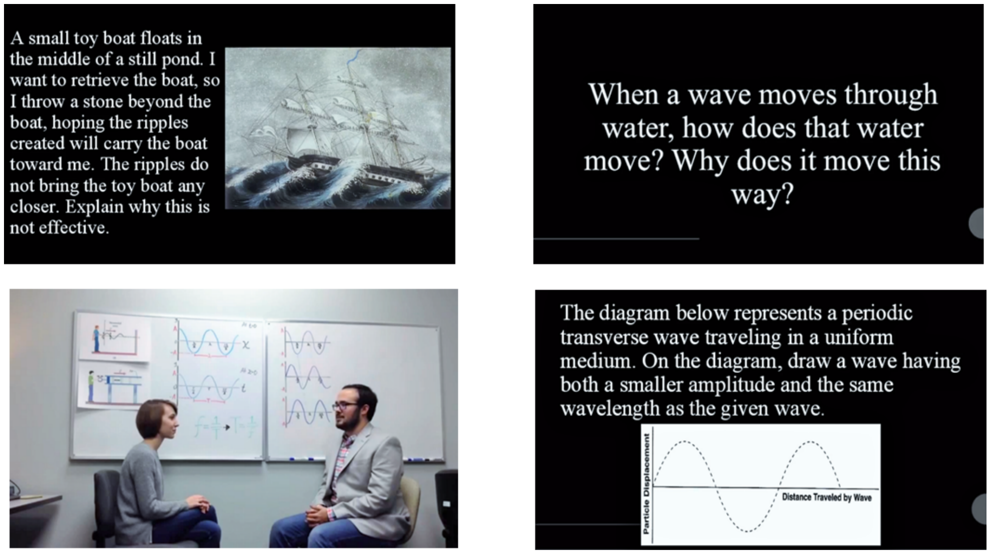

Figure 2. Screenshots of physics video and interventions

The physics content scripts were developed by a physics professor at the university where the research was conducted. In the Breakdown Scenario condition, participants were presented a scenario that did not work as expected. For example, *I have two strings of equal length. One is thinner and lighter; the other is thicker and heavier. I tie them together, end to end, and attach the thick end to a stationary pole. I grab the thin end and pull the string taut. Now I start to move my hand up and down at a constant frequency, and waves travel down the string. The wave speed is higher on the thin string and lower on the thick string. You notice that there seem to be fewer crests on the thick string at any given time than on the thin string. What is wrong with this picture?* The participant is then asked to explain the rationale (i.e., the physics principle) for why the scenario is not working.

In the Deep Questions condition, participants were given questions that could only be answered by having a conceptual understanding of the material being covered (Craig et al., 2012). For example, *When a wave moves through water, how does that water move? Why does it move this way?* The questions were classified as deep according to an existing question taxonomy (Graesser & Person, 1994; Graesser, Person, & Magliano, 1995).

The Intra-testing condition required students to complete a mathematically based physics problem. See Figure 3 for an example of an intra-testing question.

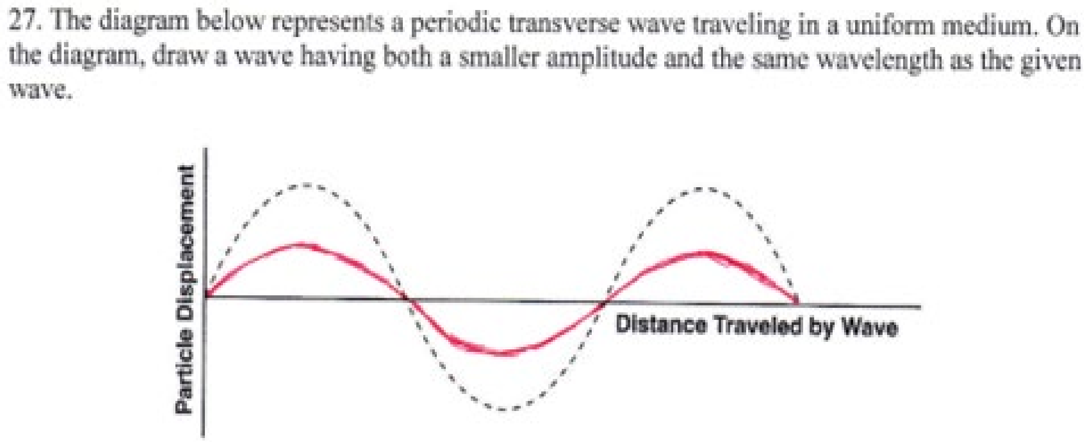

Figure 3. Example question from intra-testing condition.

The control condition was an information delivery of the content comparable to a typical physics lecture. In order to control for time on task, participants in the control condition were asked to reflect on the information that had been presented. This time of reflection occurred at the same time and was the same length as the interventions in the three confusion induction conditions.

Because the purpose of the experiment was to intentionally induce a state of confusion in the participants, all confusion induction methods were introduced before the participants had been presented the necessary information to arrive at a solution. For all conditions (i.e., breakdown scenarios, intra-testing, and deep questions), the learner was provided an answer to the manipulation (confusion induction) following the continuation of the multimedia video immediately following a predetermined time (~4 minutes). The four minutes was chosen in order to give participants time to attempt to solve the problem or question they received as part of the manipulation. The answer content was the same across all conditions. The participants were instructed to continue working on the problem even if they felt that have arrived at a solution.

Following the learning intervention, participants completed the physics posttest (counterbalanced with the pretest) followed by the last individual differences test (Achievement Emotions Questionnaire). According to Pekrun, academic emotions are emotions tied directly to achievement activities or achievement outcomes (Pekrun, Frenzel, Goetz, & Perry, 2007). The Achievement Emotions Questionnaire is a multidimensional 24-scale instrument assessing students' class-related, learning-related, and exam-related achievement emotions (Pekrun, Goetz, & Perry, 2005). Emotions measured within these categories include enjoyment, hope, pride, relief, anger, anxiety, shame, hopelessness, and boredom (alpha > .80 for 20 of the 24 scales). This scale will allow for the assessment of students' general emotional profile during a variety of learning activities. For example, a student who frequently experiences enjoyment and hope during learning may respond differently to confusion induction methods than a student who frequently experiences anxiety and hopelessness.

Results

Trained coders went back through each of the 178 video recording of each participant and noted when the participant started mind-wandering during the instructional video (i.e., stopped paying attention). Each participant video was coded by two coders to establish reliability. Any disagreements between the two raters were resolved through discussion. Based on this analysis, data were then divided into "high mind-wandering" and "low mind-wandering". High mind-wandering participants were those that spent more than 5% of the total time of the instructional video not paying attention. High mind-wandering participants are excluded from any subsequent analyses. Additionally, because all but two of the participants were classified as having a growth mindset (as measured by Dweck's Mindset Questionnaire) this test of individual difference was excluded from any subsequent analyses.

RQ1: *What are the most efficient methods of inducing productive confusion in order to increase STEM learning gains?*

A one-way analysis of variance showed a marginally significant difference in change scores (posttest-pretest) as a function of condition, $F(3,98) = 1.796$, $p = .07$ (one tailed), η^2=.052 (see Table 1 for a full list of means).

Table 1.
Change Score Means and Standard Deviations Across the Four Conditions

	Means	**Standard Deviation**
Breakdown Scenarios	2.55	2.89
Deep Questions	1.32	2.98
Intra-testing	.80	3.48
Control	.82	2.61

As can be seen from Table 1, a sizable difference existed between learning gains for those participants in the Breakdown Scenarios and the participants in the Control condition. Based on these observations, subsequent analyses excluded Intra-testing and Deep Questions.

An independent samples t-test was conducted only on the participants in the Breakdown Scenario and Control. Results revealed a significant difference between participants in the Breakdown Scenarios (M = 2.55, SD = 2.89) and the participants in the Control condition (M = .82, SD = 2.61), t (55) = 2.316, p = .02, d = .63 (See Figure 4).

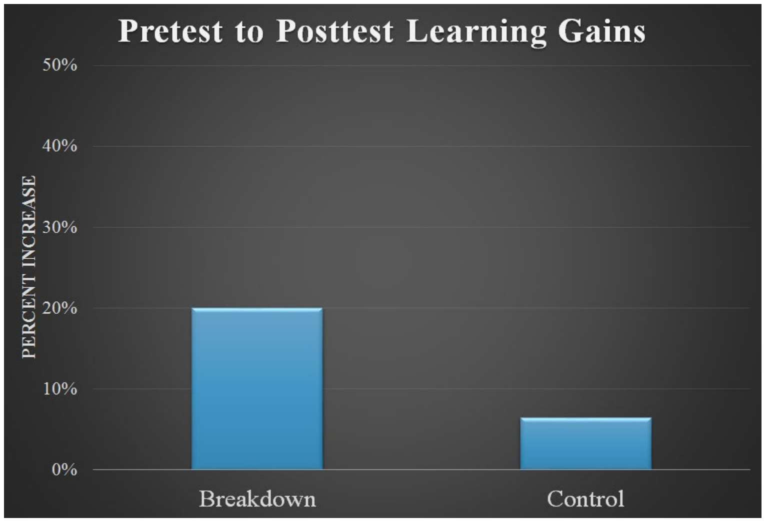

Figure 4. Learning gains as a function of condition.

RQ2: *What role and to what extent do individual differences play in regards to confusion induction and learning of complex science topics?*

To examine if any interactions existed between Attributional Complexity and Condition, a two way between subjects ANOVA was conducted. Results revealed no significant overall interaction, F (1,53) = .601, p = .442. Although there was no an overall significance between the two variables, follow up tests did reveal a significant pairwise comparison between Condition and Attributional Complexity. More specifically, participants with high Attributional Complexity in the Breakdown condition (M = 3.00) had significantly higher pretest to posttest change scores compared to those with high attributional complexity in the Control condition (M = .941), p = .03 (See Figure 5).

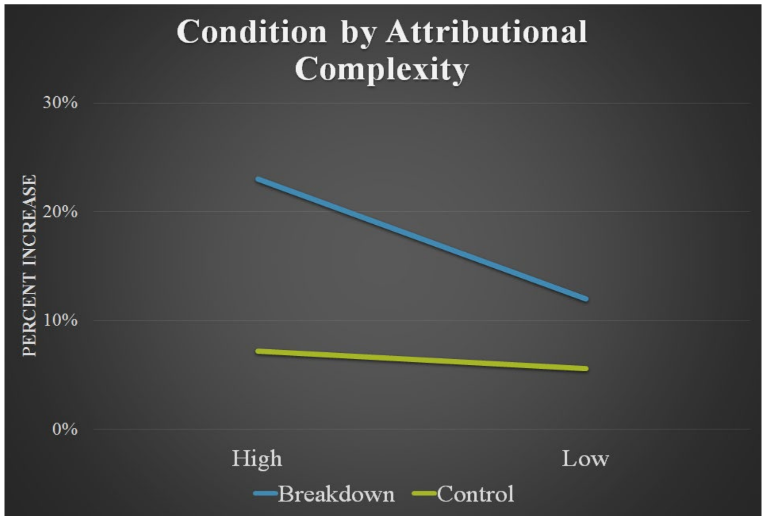

Figure 5. Learning gains for the condition by attributional complexity interaction.

A two-way ANOVA was performed to explore any possible interactions that exist between Goal Orientation (Mastery/Performance; Approach/Avoidance) and Condition. Results revealed no significant overall interaction, F (3,36) = .282, p = .838. Although there was not an overall significance between the two variables, follow up tests did reveal multiple significant pairwise comparison between Condition and Goal Orientation. Participants in the Breakdown Scenario with Performance Approach Goal Orientation (M = 4.60) learned significantly more than participants in the Control with Performance Approach Goal Orientation (M = .917), p = .019. Additionally, participants in the Breakdown Scenario with Performance Approach Goal Orientation (M = 4.60) learned significantly more than participants in the Breakdown Scenarios with Mastery Avoidance Goal Orientation (M = 1.00), p = .05 (See Figure 6).

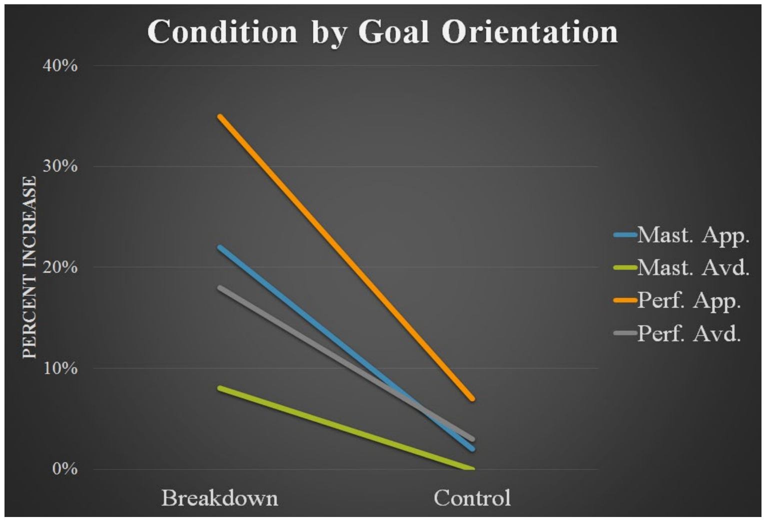

Figure 6. Learning gains for condition by goal orientation interaction.

Discussion

Based on the aforementioned results it can be seen that there does appear to be a "gold standard" when attempting to induce a state of confusion that leads to significant learning gains when compared to a traditional lecture based format. *Breakdown Scenarios* led to significantly higher change scores compared to the control condition (i.e., lecture). No significant differences were found among the intra-testing condition, deep questions condition, and the control condition. However, given the significant pairwise comparisons that were discovered between Breakdown Scenarios and certain individual difference measures, it appears that Breakdown Scenarios are better suited for certain subsets of students. The results from the current study begin to shed light on the learner profile that benefits the most from confusion induction. More specifically, *Breakdown Scenarios* appear to be most beneficial for those learners who are either: 1) high in Attributional Complexity or 2) have a Performance Approach goal orientation.

It is the opinion of the authors that a student high in attributional complexity would be more likely to consider multiple reasons for why a phenomenon would occur (in general). Because of this, the student might be more open to comparing multiple perspectives which in turn would help them resolve their confusion because they'd be willing to actively work through not only why one perspective was correct but also why another perspective was incorrect (reaching a deeper understanding and leading to learning).

Achievement goal theory refers to the reasons why individuals engage in a task. The results from the current study are in concert with previous research which has found that goal orientation is related to learning and learning strategy implementation. Previous research has found, for example, that learners with mastery-approach goals (MAp) have shown positive emotions and interest, elaborative learning strategies and effective self-regulation, and help-seeking and cooperativeness. However, the research on performance goals is less conclusive. For example, Ames, Maehr, Fisher, Archer, and Hall (1989) discovered that characteristics such as anxiety, disorganization, superficial learning habits, and low exam performance were associated with students who exhibit performance goals whereas other research has found that students with performance approach goals actually outperform students with mastery goals (Schunk, Meece, & Pintrich, 2014). In terms of confusion induction and learning, the results can be explained by the following: 1) students completed these experiments in a social setting (i.e., in the presence of the experimenter) which in turn caused perception to be a factor. Because of this, students who enjoy demonstrating their competence to others benefited. Additionally, students who were worried about looking incompetent may generally not have enjoyed the instances of confusion (regardless of setting) because an instance of confusion inherently meant that there was a moment of uncertainty about one’s current knowledge. Because of this, students with avoidance goal orientation benefited the least from the Breakdown Scenarios (i.e., confusion induction).

In a traditional learning environment, instructors are responsible for adapting to both the cognitive and affective states of learners as they face challenges and work to maintain motivation. However, the educational landscape is shifting to a more non-traditional format. Over a quarter of students (28%) have taken at least one online class and that number has been consistently growing (Allen, Seaman, Poulin & Straut, 2016). The internet (e.g., Google) is beginning to become the "go-to" source for information (Roscoe, Grebitus, O'Brian, Johnson, & Kula, 2016). Although this trend of moving away from human to human interaction has altered the traditional educational landscape, students still experience the same types of positive and negative emotions during learning (Calvo & D'Mello, 2011; Meyer & Turner, 2006; Pekrun, 2010). The anticipated findings from the current study could have a potentially significant impact on both traditional and online education. Regardless of the presentation medium, productive confusion, an instance of uncertainty during learning that is successfully resolved, is one emotion that is essential for successful learning (Craig, 2012a; D'Mello & Graesser, 2012). If strategies to induce confusion that promote successful resolution could be identified and implemented with minimal effort, these could be widely beneficial for college courses, junior/high school lesson plans, e-textbooks, and online course content.

It warrants mentioning that the authors are not arguing that *all* confusion works *all* of the time. It is the belief of the authors that confusion alone is not the catalyst to learning but the underlying cognitive mechanisms activated by the confusion are what lead to higher learning. As shown from the results, productive confusion does have the potential to propel some learners into a higher altitude of understanding, but for others it may produce negligible or perhaps even detrimental learning gains. The difficulty of course is attempting to identify the "zone of optimal confusion" for each individual learner. Although the current study did explore how certain individual differences impact confusion and learning (i.e., attributional complexity and goal orientation) more work needs to be done in this area to understand when confusion works and for who. Additionally, although these results are exciting and immediately applicable, caution should be taken in generalizing these findings to other domains. It is suspected that certain disciplines are better suited for the introduction of desirable confusion but further research is needed in this area.

This desirable confusion paper sought to advance student learning of STEM disciplines. It is the hope of the authors that we have discovered a new supplemental method that is more inclusive and motivating than traditional STEM instructional approaches.

References

Allen, I.A., Seaman, J., Poulin, R., Straut, T.T. (2016). Online report card V- tracking online education in the United States. *Babson Survey Research Group and Quahog Research Group*.

Ames, C., Maehr, M. L., Fisher, A., Archer, J., & Hall, H. (1989). Achievement goals and the structure of classroom learning. Paper presented at the annual meeting of the *American Educational Research Association*, San Francisco.

Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C. (2010) Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68 (4), 223-241.

Brown, J. S., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 4, 379-426.

Calvo, R. A., & D'Mello, S. K. (2011). *New perspectives on affect and learning technologies* (Vol. 3). New York: Springer.

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.

Craig, S. D., Gholson, B., Brittingham, J. K., Williams, J. L., & Shubeck, K. T. (2012). Promoting vicarious learning of physics using deep questions with explanations. *Computers & Education, 58*(4), 1042–1048.

D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*, 16, 3-28.

D'Mello, S. K., Craig, S. D., Witherspoon, A. W., McDaniel, B. T., & Graesser, A. C. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction, 18*(1-2), 45-80.

D'Mello, S. K., & Graesser, A. C. (2011). The half-life of cognitive-affective states during complex learning. *Cognition and Emotion*, 25(7), 1299-1308.

D'Mello, S. K., & Graesser, A. C. (2012). Dynamics of affective states during complex learning, *Learning and Instruction*, 22, 145-157.

D'Mello, S. K. & Graesser A. C. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios, *Acta Psychologica, 151*, 106-116.

D'Mello, S., & Millis, C. (2014). Emotions while writing about emotional and non-emotional topics. *Motivation and Emotion, 38*(1), 140-156.

D'Mello, S. K., Lehman, B. A., Pekrun, R., & Graesser, A. C. (2014). Confusion can be beneficial for learning, *Learning & Instruction*, 29(1), 153-170.

D'Mello, S. K., Lehman, B. A., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education*, 20(4), 361-389.

Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501-519.

Fletcher, E J. O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). *Attributional complexity: An individual difference measure. Journal of Personality and Social Psychology, 51*, 875-884.

Graesser, A. C., Chipman, P., King, B., McDaniel, B., & D'Mello, S. K. (2007). Emotions and learning with AutoTutor. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (pp. 569-571). Amsterdam: IOS Press.

Graesser, A. C., & McMahen, C. L. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology, 85*(1), 136-151.

Graesser, A. C., & 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.

Graesser, A., & Person, N., (1994). Question asking during tutoring. *American Educational Research Journal, 31 (1)*,104–137.

Graesser, A. C., Pearson, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*, 495- 522

Lee, D.M., Rodrigo, M.M., Baker, R.S.J.d., Sugay, J., Coronel, A. (2011) Exploring the relationship between novice programmer confusion and achievement. *Proceedings of the 4th bi-annual International Conference on Affective Computing and Intelligent Interaction*.

Lehman, B., D'Mello, S., & Graesser, A. (2012). Interventions to regulate confusion during learning. In. S. A. Cerri, & B. Clancey (Eds.), *Proceedings of 11th International Conference on Intelligent Tutoring Systems* (ITS 2012) (pp. 579-581). Berlin: Springer-Verlag.

Lehman, B., D'Mello, S. K., Strain, A., Mills, C., Gross, M., Dobbins, A., Wallace, P., Millis, K., & Graesser, A. C. (2013). Inducing and tracking confusion with contradictions during complex learning. *International Journal of Artificial Intelligence Special Issue: Best of AIED* 2011, 22(2), 71-93.

Lehman, B. A., Matthews, M., D'Mello, S. K., and Person, N. (2008). What are you feeling? Investigating student affective states during expert human tutoring sessions. In B. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems* (pp. 50-59). Berlin, Heidelberg: Springer-Verlag.

Lehman, B., D'Mello, S., & Graesser, A. (2013). Who benefits from confusion during learning? An individual differences cluster analysis. In K. Yacef, C. Lane, J. Mostow, & P. Pavlik (Eds.), *Proceedings of 16th International Conference on Artificial Intelligence in Education (AIED2013)* (pp. 51-60). Berlin Heidelberg: Springer-Verlag.

Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, 11, 357–380.

Muller, D. (2012). The key to effective educational science videos [Video file]. Retrived from: https://www.youtube.com/watch?v=RQaW2bFieo8

Meyer, D.K., Turner, J.C. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, 18 (4), 377-390.

Pekrun, R. (2010). Academic emotions. In T. Urdan (Ed.), *APA educational psychology handbook* (Vol. 2). Washington, DC: American Psychological Association.

Pekrun, R., Frenzel, A., Goetz, T., & Perry, R. P. (2007). The control- value theory of achievement emotions: An integrative approach to emotions in education . In P. A Schutz & R. Pekrun (Eds.), *Emotion in education* (pp.13- 36). San Diego, CA: Academic Press.

Pekrun, R., Goetz, T., & Perry, R. P. (2005*). Achievement Emotions Questionnaire (AEQ). User's manual*. Munich, Germany: Department of Psychology, University of Munich.

Piaget, J. (1952). *The origins of intelligence*. New York: International University Press.

Rodrigo, M.M.T. & Baker, R.S.J.d. (2011) Comparing learners' affect while using an intelligent tutor and an educational game. *Research and Practice in Technology Enhanced Learning*, 6 (1), 43-66.

Roscoe, R. D., Grebitus, C., O'Brian, J., Johnson, A. C., & Kula, I. (2016). Online information search and decision making: Effects of web search stance. *Computers in Human Behavior*, *56*, 103-118.

Schunk, D.H., Meece, J.R., Pintrich, P.R. (2014). *Motivation in education: Theory, research, and applications* (4th ed.). Boston, MA: Pearson.

Sullins, J., & Goza, K. (2013, March). Exploring factors that affect question quality recognition and question generation*. Paper presented at the annual meeting of the Southeastern Psychological Association*, Atlanta, Georgia.

Sullins, J., Goza, K., Smith, H., Moore, A., & Morrow, J. (2014). The use of anomalous information to improve student question generation in a multimedia learning environment**.** *Proceedings of the 27th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*. Pensacola, FL: AAAI Press.

Sullins, J., & Graesser, A. C. (2014). The relationship between cognitive disequilibrium, emotions, and individual differences on student question generation. *International Journal of Learning Technology*, *9*(3), 221-247.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Bagget, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21, 209–249.