I doubt, therefore I learn

how to use confusion to foster deeper learning and students' confidence in online learning environments

Mémoire réalisé en vue de l'obtention de la Maîtrise universitaire

en sciences et technologies de l'apprentissage et de la formation

Djamileh Aminian
Genève, le 9 mai 2022

Direction

Prof. Gaëlle Molinari

Jury

Dr. Catherine Audrin

Prof. David Sander

Abstract

The purpose of this research was to examine the influence of a confusion-based instructional design in online learning environments on learning outcomes and learners' confidence. This proposition of instructional design is an implantation of D'Mello and Graesser's Zone of Optimal Confusion (2012). This model offers a theoretical framework to optimize the learning process by inducing a type of confusion that fosters deeper and more sustainable learning. We recruited university students to follow an e-learning course on First Aid. Participants were tested on knowledge before and after following the course, either in a classic direct instruction or in a confusion-based design. They also reported on emotions and an Error Orientation Questionnaire to determine what factors influence the response a specific design. Despite strong learning gains in both conditions, results showed no significant differences between the two designs. Suggestions for further research are discussed.

Keywords

Epistemic emotions; Confusion; Learning strategies; Instructional Design; Error Orientation Style

Table of Content

1. INTRODUCTION 4
   - 1.1 Introduction 4
   - 1.2 Context 5
   - 1.3 Scope of this work 7
2. A REVIEW OF RELEVANT THEORETICAL MODELS 9
   - 2.1 The cognitive side of learning 9
   - 2.2 The affective side of learning 11
   - 2.3 Digital-based Learning 24
3. FOCUS OF THE RESEARCH 27
   - 3.1 Research question and hypotheses 28
4. METHOD 29
   - 4.1 Participants 29
   - 4.2 Material 30
   - 4.3 Experimental procedure 38
   - 4.4 Experimental design 40
   - 4.5 Statistical Analysis 42
5. RESULTS 43
   - 5.1 Descriptive analysis 43
   - 5.2 Effects on Learning gains 45
   - 5.3 Effect on Confidence 46
   - 5.4 Effects on Emotions 47
   - 5.5 Correlations between variables 48
6. DISCUSSION 50
   - 6.1 Hypotheses and results 50
   - 6.2 Limits 52
   - 6.3 Perspectives 54
7. ACKNOWLEDGMENTS 56
8. BIBLIOGRAPHY 57
9. APPENDIX A - COURSE CONTENT 64
10. APPENDIX B - QUESTIONS 65
11. APPENDIX C - QUESTIONNAIRES 66

1. Introduction

“Certainty was to curiosity what the sun was to the wings of Icarus.

Where one shone forcefully, the other couldn't survive.”

Elif Shafak, Three daughters of Eve, 2016

1.1 Introduction

Like Arthur in his quest for the Holy Grail, educators of all persuasions have been looking for the best methods of teaching to foster the best learning outcomes. School teaching, academic curricula, professional training, all these fields have, in the past decades, adapted to new trends and tools, as ideologies evolved, and technologies created new affordances and made previously unthinkable practices possible.

With a focal point on cognitive processes in the 20th century, research in education has since then included the study of affective states to understand learning. Questions were raised on what emotional states arise, in what conditions, for what type of learners, with what influence on the learning outcomes, and many more.

Through the advent of positive psychology and the self-help industry, and maybe because so many people report some form of trauma from their school years, from being bored and disengaged to feeling in constant distress, a lot of modern educators seem to assume that learning is optimal while feeling positive emotions, like flow (Csikszentmihalyi, 1990) or interest (Silvia, 2008).

How counterintuitive therefore to recommend negative emotions to foster a deeper and more sustainable learning! Yet, in this age of “quick and easy” micro-learning, can we really learn effectively if none of our ideas are challenged? if no phase of the process requires some form of effort? without feeling a little bit of discomfort?

Broadly speaking, learning entails acquiring new information, comparing it to existing mental models and deciding - more or less consciously - if it should be integrated or ignored. When this information is sufficiently relevant for the individual and distinct from the current models, it raises doubt. Doubt allows for critical thinking; it signals to the brain that some form of deep thinking is required. Doubt asks for attention and working memory resources that may, under certain conditions, initiate a shift from the quick and instinctive System 1 to the slower, more effortful, and focused System 2, to use Daniel Kahneman's popular model (Kahneman, 2011).

Doubt opens doors to change, whereas belief rejects the new to protect the old. Therefore, we posit that doubt is a vital and essential part of learning and that instructional design models should consider how to best raise doubt in learners to guide them towards stronger learning outcomes.

This work aims to study how doubt can be induced and fostered, and under what conditions it is beneficial to the learning process. As more people are training with computer-mediated applications, we will specifically focus on understanding how to integrate doubt in digital learning to foster retention and depth of knowledge.

1.2 Context

Could anyone live without learning? Learning is critical to the survival of human beings: acquiring new information, understanding it, storing it and being able to use it in the future underlies evolution. Yet, even after millennia of pondering how we learn and how to teach in the most optimal way, we still struggle to understand exactly what goes on during the process of learning. Faced with new or conflicting information to existing mental models, what factors influence the choice, conscious or not, to adapt the models and thus learn?

Research in Education has long focused on the cognitive or social mechanisms that contributed to the learning process. Surprisingly, how emotions influence these mechanisms have been mostly left out until the end of the 20th century. Since then, a growing body of research

(Linnenbrink-Garcia & Pekrun, 2011; Pekrun, Goetz, Titz & Perry, 2002) has emerged to study emotions in educational settings. These studies focus on what are called *epistemic emotions*, that is, emotions that occur while learning, or "emotions that arise when the object of their focus is on knowledge and knowing" (Muis, Psaradellis, Lajoie, Di Leo & Chevrier, 2015, p.173).

Major studies on epistemic emotions include slightly different sets of emotions, still most of them agree on interest, curiosity, surprise, confusion, boredom, frustration, and anxiety (Muis et al., 2015). Other researchers talk about "learning-centred emotions", adding engagement/flow and happiness to the previous list (Rodrigo, Mercedes & Baker, 2011). These studies investigate the influence of these emotions on learning in general, from knowledge retention to learners' engagement, how and when they occur, how beneficial or detrimental they can be to learning, and in what conditions.

Whereas interest and curiosity seem to be positively correlated with learning (Muis et al., 2015), and boredom and frustration negatively (Pekrun & Stephens, 2012), confusion has a more ambivalent relationship to learning: it can predict either a positive or a negative outcome, as it depends on a learner's ability to resolve it (Craig, Graesser, Sullins & Gholson, 2004).
Confusion emerges with doubt, that is, when an individual is unable to make sense of a discrepant information, triggering cognitive imbalance or *incongruity*. It persists until the incongruity is resolved (D'Mello & Graesser, 2012; D'Mello, Lehman, Pekrun & Graesser, 2014). Confusion seems thus to be a normal and regular part of the learning process, even welcomed, as its presence tends to encourage the questioning of old models and their adjustment with new, relevant information.

But feeling confused has complex implications: it gives a conscious signal that the system is out of balance, that there is a disequilibrium in the homeostatic state of the organism. As our bodies constantly adapt to the change of internal or external elements, so, it would seem, do our minds need to reach a state of equilibrium that allows us to function. But what is learning if not the constant processing of new information and its accommodation with existing models of the world? When is the mind, in its cognitive and affective dimensions, really at rest, at an equilibrium point? Can change, adaptation, and therefore learning happen, in a state of pure balance? Imbalance - created by the new, the unexpected, the challenging - drives the growth and development of an individual.

Humans have developed sophisticated mostly unconscious mechanisms to process information. With limited resources in terms of working memory, attention, time, and energy, we developed mechanisms to choose, filter, select. That is why imbalance cannot be considered as a pure antecedent to learning: if we cannot, or will not, devote cognitive resources, we will simply discard the discrepant information and keep the models untouched.

How is this choice made? *Appraisal theories* (e.g., Lazarus, 1991) greatly help us understand what happens during learning. Emotions are first triggered by the evaluation of *relevance*: when faced with a stimulus, the system evaluates the need to allocate resources to its processing, based on its connection to our goals or needs. It answers the question: “Are there good enough reasons for me to take this into account?”. This is very interesting for learning, as it highlights that appraisal is a dynamic process, which depends on multiple competing factors, and that relevance does not only vary from one person to another (our goals and needs differ), but also differs based on availability of resources, e.g., fatigue, health, current concerns, stronger competing stimuli, current affective states, to name a few.

A second level of appraisal, hugely relevant to learning, is the assessment of one’s ability to face the new information. To the question “Am I able to do it?”, a negative answer may trigger anxiety, leading to disengagement. Again, detecting the outcomes of these appraisals is of the highest importance to instructional designers, as an adequate level of support is necessary to sustain learning, and this support needs to respond to an ever-changing dynamic appraisal of the learning in progress. Assisting learners in their journey by measuring emotional states and giving appropriate support seems to be one of the major current challenges in education today.

1.3 Scope of this work

Amongst the numerous definitions for learning, one strikes us as particularly relevant to this work: *learning means taking a risk*, that is, to learn is to accept change; to learn is to admit that what we see as truth is only a fragile model of the world, constantly subject to challenging information from our environments, requiring adaptation in order to survive.

Paradoxically enough, although change is one of the only certainties one can have about the world, human beings are often reluctant to accept new information that threaten the equilibrium of their model. Nonetheless, doubt, incongruity and their emotional marker, confusion, help us to be curious, to learn, to accept new information. Ignoring confusion is risky. Without doubt, without accepting the possibility that what we know is a simplification of our reality - or just plain wrong -, it is not possible to evolve. A person's relationship to doubt and error seem therefore particularly interesting to factor in when designing with confusion. Quoting Carol Dweck's Mindset theory, learners' error orientations and mindsets can have a significant effect on their performance: learners with *growth mindset*, i.e., who perceive errors and failure as an opportunity to learn and grow perform better in a class than learners with a *fixed mindset*, who think intelligence is fixed and errors are threatening to their self-esteem (Dweck, 2006).

Social, physiological, mental, and emotional factors are all influencing the way we learn at a specific moment, as they are dynamic states. Considering all these parameters when designing a course is fundamental, yet unrealistically ambitious. In this work, we leave out the social and physiological factors to focus primarily on the effect of confusion on mental processes involved in learning.

We will thus explore the role of confusion in the design of e-learning courses that promote stronger student engagement and deeper learning and determine if confusion-based instructional design can benefit learners, specifically in the light of their error orientation profile. The next section presents a synthesis of relevant models and theories. It is followed by an experimental study on the influence of a confusion-based design for an online e-learning module and its effects on learning outcomes. We will then discuss the results and offer considerations to explore this subject further

2. A review of relevant theoretical models

2.1 The cognitive side of learning

Countless definitions have been given to the act of learning over the years. Influenced by historical events, cultural and political ideologies and technological developments, experts of their time have approached learning and teaching with multiple bias.

Most modern educational models are based on constructivism and socio-constructivism, while integrating methods like gamification that relies heavily on behaviourism concepts, or insights from cognitive neuroscience, bringing an essential understanding of how humans learn through perception, attention, and memory. We ground our research in constructivism, the leading framework in today's instructional design, giving a strong theoretical basis in which to study epistemic emotions, particularly confusion. Socio-constructivism is also defined, as it offers an insightful perspective on the support to help regulate confusion.

2.1.1 Constructivism: the learner as the agent of their development

Constructivist theorists, led by Jean Piaget, propose a vision of learning as the construction of one's own knowledge, driven by the fundamental need to give meaning to one's experiences. This stresses the individualisation of learning, emphasising that all information from the environment is compared with prior internal representations and that new knowledge is built around these two poles. Knowledge can therefore only be personal, and the individual plays an active role: constructivism sees learning as a permanent reconstruction of one's reality. As such, learning is strongly conditioned by the individual's goals and needs, i.e., their motivation. If new information is not considered relevant and no motivation is triggered, it is unlikely that any learning will take place, thus consistent with Appraisal Theories.

Learning is seen as a cognitive activity, consisting of constructing one's own knowledge, based on previous knowledge and new information from the environment. Learners restructure their mental models as they acquire new knowledge. Piaget compares them to scientific researchers formulating hypotheses about the world to test them. The imbalance created by the comparison between previous knowledge and new knowledge must be balanced by a process of integration. This cognitive incongruity is called *Cognitive Conflict*. To constructivists, for learning to occur, this *conflict*, that is the incongruity, must be resolved. This can be done either through learning (integrating or adjusting the models) or other regulation strategies, e.g., discarding discrepant information. This emphasises how learners may react in different ways toward confusion.

In a constructivist pedagogy, teachers seek to emancipate the learners, giving them the tools and autonomy to construct their own knowledge. This mainly leads to setting up conditions that favour the emergence of a *cognitive conflict*, i.e., giving the learner the space and curiosity to compare their models with new information and to go through a process of analysis, to reject, integrate partially or totally the new information. Confusion is welcomed and encouraged. In confusion-based instructional design, this means assessing current knowledge and providing sufficiently complex activities to the learner to get confused but being able to regulate it with no or minimal support. How to define the type and level of support to give a learner is guided by the educational approach we present, socio-constructivism.

2.1.2 Socio-constructivism: Learning through and with others

Parallel to Piaget's work, other researchers - among which popular figures like Lev Vygotsky, Maria Montessori and Jerome Bruner - have suggested a form of constructivism that includes the social dimension as a major component in the learning process, creating socio-constructivism.

Vygotsky's Zone of Proximal Development (ZPD) is of particular interest for our research. He defines two thresholds to be identified in order to measure a learner's progress: their current knowledge as a starting point and their developmental potential, i.e., what they can achieve with the help of an expert. The ZPD represents the temporal difference between what the learner has achieved on their own and what they can achieve with appropriate scaffolding: if initially, a learner benefits from strong scaffolding to guide them in solving the problem, the more the learner shows their ability to carry out the activity independently, the less scaffolding will be provided, to the point where it will disappear.

This process can be seen as an adaptation of the cognitive conflict with a social dimension. This *socio-cognitive conflict* relates to the use of disruption and incongruity as constituent elements of learning. Teaching using a socio-constructivist approach implies setting a proper learning environment for the student to compare their internal models to the information gathered from external sources, within a specific dynamic range - or zone - where the disruption is a support for learning.

By applying this concept of Zone of Proximal Development to emotions, learners must be kept in a specific range, where they are sufficiently challenged not to be bored, and not overwhelmed to become anxious and discouraged. This dynamic process is complex and calls for regular monitoring to offer the right amount of assistance, thus controlling persistent frustration or confusion. This concept of fragile equilibrium and adaptive scaffolding has been a major influence for D'Mello and Graesser's *Zone of Optimal Confusion* (D'Mello & Graesser, 2012). This model of epistemic emotions centred around confusion is one of the main sources for our work, and is presented in the next section, where we define the major models and theories for the study of confusion.

2.2 The affective side of learning

Numerous research has been conducted on *epistemic emotions*, showing that emotions profoundly relate to the act of learning (Vogl, Pekrun, Murayama, Loderer & Schubert, 2019) and are not only relevant, but crucial to learning (Stein & Levine, 1991; Kort, Reilly & Picard, 2001). While a lot of studies focus on positive emotion, like *interest* (Silvia, 2008) or *flow* (Csikszentmihalyi, 1990) a growing body of research has been directed on more ambivalent emotions like confusion. As already mentioned previously, this emotion is considered to be the emotional marker of cognitive imbalance and detecting its occurrence to help learners regulate it seems to be predictive of deeper and more sustainable learning (D'Mello & Graesser, 2012).

When learners experience cognitive conflict, they may initially first feel surprise, that may be followed by curiosity and/or confusion, depending on the level of discrepancy of the incongruity and the learner's resources to resolve it (Muis et al., 2015).

However, not all learners respond to confusion in the same way (Long, Luo, Gao & Hu, 2019; Lehman, D'Mello & Graesser, 2013; D'Mello et al., 2014). According to one study (D'Mello et al., 2014), confusion can be *productive* if it can be regulated and lead to deeper learning, whereas *persistent confusion* is a predictor of frustration and disengagement (see figure 1).

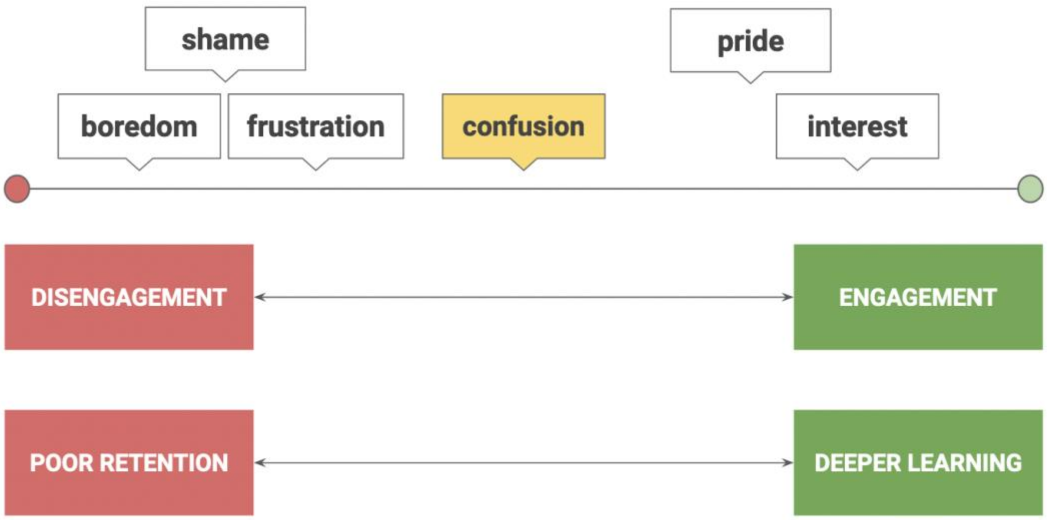


Figure 1. Academic and epistemic emotions model, based on D'Mello, Lehman, Pekrun & Graesser, 2014

Interestingly, parallel research in education with a similar orientation, like Productive Failure (Kapur, 2016; Loibl, Roll & Rummel, 2017) have studied how fostering failure helps generate stronger conceptual learning in students. These authors also point out that there are numerous contextual and individual parameters that influence its success, stemming from cognitive, affective, social and cultural dimensions (Sinha & Kapur, 2021). They provide an interesting list of parameters that we have used in our study to control confusion-induction, which will be detailed in the section on method.

As distance learning becomes more and more important in adult education, the question arises more fundamentally for e-learning environments: how to provide the right kind of support, personalised to the needs of the student, when they learn alone? This type of *scaffolding*, as theorised by Jerome Brunner (1975), is crucial when discussing confusion, as it is not the emergence but the regulation of the emotion that is conducive to learning (D'Mello & Graesser, 2012).

To understand how emotions arise and can be regulated during learning, we will first give an overview of the current state of the art on epistemic emotions, describing major models and theories. Then we will detail two important models for the study of confusion.

2.2.1 Emotions in Education & Instructional Design

It is noteworthy at this stage to discuss the existence of an ongoing debate on the status of confusion as an emotion. For D'Mello and Graesser (2012), confusion is considered an emotion. Some researchers nuance the use of the terms *emotions* for epistemic emotions, arguing that we should be referring to *cognitive-affective states*, that is a "blend of affect and cognition" (Baker, D'Mello, Rodrigo & Graesser, 2010). For a review on the debate, see Linnebrink and Pekrun, 2011.

In this work, as we base our study primarily on D'Mello and Graesser "Zone of Optimal Confusion" (2012), we will refer to confusion as an emotion, as proposed by these authors. We postulate that the "persistent" aspect of emotion described by these authors may refer to the recurrent occurrences of confusion episodes, on the same topic, without major occurrences of other types of emotions.

As Craig et al. (2004) point out in their article, the models for emotions developed by psychologists since the mid-1970's, e.g., Paul Eckman's model with six basic emotions, were not particularly useful to discuss affective states during learning, as fear or disgust are unlikely to occur during learning. Researchers in education thus set out to identify which emotions appear during the learning process and which ones are beneficial or detrimental to it. Theories from other fields in psychology, like motivation or goal fixation, gave a strong theoretical framework to detect how and which emotions were relevant for learning.

Craig et al. (2004) explored the role of six emotions on learning: frustration, boredom, flow, confusion, eureka and neutral. They found significant correlations between learning outcomes and boredom, flow, and confusion, with a positive correlation between confusion and learning, consistent with the constructivist theory that cognitive disequilibrium is a precursor to deep learning. Results also showed, as expected, a negative correlation with boredom and a positive correlation with flow, consistent with Csikszentmihalyi's theory of Flow (Csikszentmihalyi, 1990).

Stein and Levine (1990) state that emotional experience is almost always associated with attending to and making sense out of incoming information. This is consistent with appraisal models proposed by Lazarus (1984) or Scherer (2009), where novel information is tested for relevance and processed accordingly. If relevance is found, the autonomic nervous system is activated, and the emotion produced. For these authors, it would mean that learning takes place if an emotional episode occurs. This sheds light on two important aspects for our study: (1) that the relevance of external information and the subsequent emotional response is unique to an individual (and, even further, to an individual *in a specific context at a specific moment*), and (2), that, in a sense, to foster learning, educators must ensure, or at least allow the emergence of a certain level of stress in their instructional design. Here, we use the word *stress* as the occurrence of an event that is relevant enough and with a degree of arousal sufficient to trigger its processing and resolution by the system.

2.2.2 Confusion

Studies on confusion can be traced back to Darwin and his research on frowning. Confusion appeared during concentration and intense thinking, often accompanied by feelings of mental imbalance (Piaget, 1952) and cognitive dissonance (Festinger, 1957). A current definition of confusion is that it occurs when anomalies or disruptions arise (conceptual novelty, contradiction, something unexpected) in the process of understanding (D'Mello & Graesser, 2013).

In a study designed to compare epistemic emotions across learning systems, irrelevant of learners' characteristics and methodology used to measure emotions (Baker, D'Mello et al, 2010), confusion was found to be the second most common state (13% of student time), after *engaged concentration* (60% of student time). Confusion was experienced when learners were faced with impasses, and was productive when it only lasted short periods, meaning it could be resolved. These findings are consistent with impasse-driven theories of learning (VanLehn, Siler, Murray, Yamauchi & Baggett, 2003).

Students who showed little motivation or had little to no prior knowledge on the topic were observed to *game the system*, that is avoiding mobilising cognitive resources (Rodrigo, Baker, Lagud, Lim, Macapanpan & Viehland, 2007). In this regard, confusion seems to be contingent on a person's motivation and knowledge level. Moreover, learners faced with confusion may give up because of their attribution styles (Weiner, 1972): they may think that they feel confused because of inherent poor abilities instead of experiencing challenges as a part of the normal learning process. This is in line with goal-achievements theories (Dweck, 2002), which we describe later in this section.

To understand the emergence of confusion and its relation to other epistemic emotions, three models are of particular interest:

1. **Kort's Affective Model of Interplay Between Emotions and Learning** (Kort et al., 2001), usually referred to as *Kort's Spiral Model*, offers a framework in which to see the dynamic flow of emotional responses in learners.
2. D'Mello and Graesser's **Zone of Optimal Confusion** (D'Mello and Graesser, 2012), already mentioned above in relation to Vygotsky's Zone of Proximal Development, sheds specific light on confusion.
3. **Productive Failure** (Kapur, 2016) creates a link between confusion, failure, and instructional design.

2.2.2.1 Kort's Spiral Model

When Kort et al. (2001) decided to work on a model to highlight the importance of emotions during learning, they were troubled by the "polished form" in which teachers present novel content to their learners. For these authors, vital elements of the learning process were missing, namely the possibility to "make mistakes, recover from them, deconstruct what went wrong, and start over again."

This insight led them to create one of the only models that link emotions to learning, describing the interaction between emotions and learning phases. The model (Figure 2) consists of a four-quadrant diagram, with the valence of emotions on the x-axis (from negative to positive) and learning outcomes on the y-axis, ranging from *constructive learning* (integrating new information) to *un-learning* (discarding misconception). Interestingly, the authors discussed adding a third axis to their model, the *knowledge* axis, that would spiral upwards, thus depicting the cyclic process students follow in building their knowledge. To our knowledge, the model has not been updated to include this third axis.

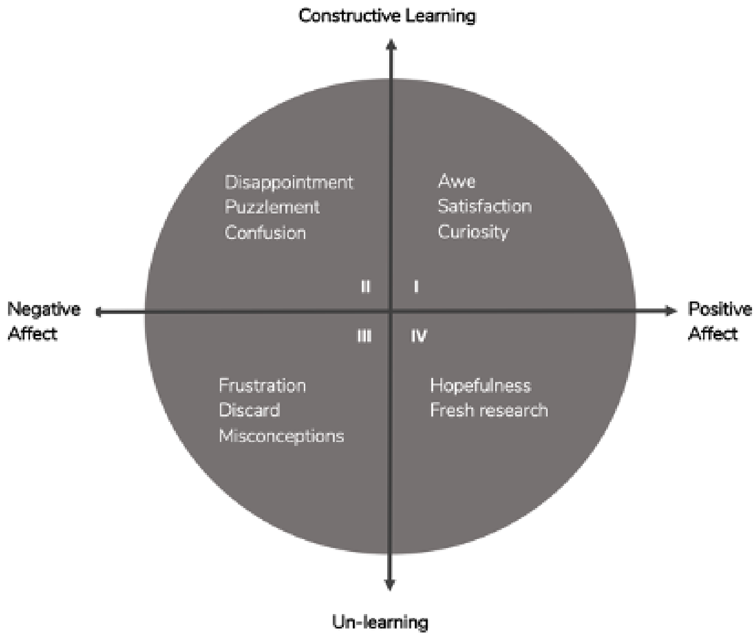


Figure 2. Kort's Affective Model of Interplay Between Emotions and Learning

In Quadrant I, a learner experiences primarily positive emotions, feeling curious about the content and looking forward to learning more, whereas, in Quadrant II, they feel puzzled by some information and experience negative emotions. The authors point out that learners can move from Quadrant I to II or start directly in Quadrant II. The upper part of the model concentrates on building new knowledge, adding new information to their existing models or comparing it with what they know. The unpleasant phase of unlearning misconceptions is shown in Quadrant III, where the need to redesign the existing model can lead to feelings of frustration if not adequately supported. Discarding misconceptions can then open new perspectives, moving back to positive emotions, in Quadrant IV.

The authors postulate that students can be in several quadrants at the same time, thus feeling several emotions in parallel: for example, a student could feel frustration (Quadrant III) by resenting the energy needed to eliminate misconceptions, while feeling interested by the new information (Quadrant I), this interest being the driver to accept the idea of discarding prior knowledge.

This model offers a dynamic perspective on emotions during the learning process. The authors argue that going through the cycle is part of the process and feeling only positive emotions is not necessarily conducive to learning. They advocate bringing failure back into the learning process instead of shielding learners from negative emotions, such as doubt, confusion, or even frustration, as they are a critical part of the process. This approach would support learning on two levels: (1) by reinstating doubt as a natural and vital part of learning, students would be less likely to give up when experiencing negative emotions, and (2) they would engage in deeper information processing, thus making learning more sustainable.

These conclusions seem robust when compared with Carol Dweck's *Mindsets* theory (2006), which states that students learn best when they are in a *growth mindset*, meaning they welcome failure as an opportunity to improve their learning strategies, as opposed to a *fixed* or *performance mindset* that causes them to protect themselves by rejecting imbalances, threatening self-esteem. Current work on Productive Failure (Kapur, 2016) strongly supports these hypotheses.

One of the main interests of this work is its applicability to digital learning environments. The authors discuss the creation of an automated learning companion that would provide adapted support after identifying the Quadrant in which the learner currently is. Based on their specific needs, the companion could provide adaptive support, from no intervention when in Q1 to progressive hints, like questions to dig deeper up to the answer itself (Q2 and Q4) and emotional support when accommodation of internal models brings discomfort (Q3). The authors argue that this emotional support is the most uncertain of all types of scaffolding that can be integrated in course designs, as learners can have drastically different emotional reactions and coping mechanisms in place.

This proposition gives us an interesting foundation to create adapted support in digital learning environments. It would require the assessment of two elements: the student's cognitive-emotive state and their progress in the learning process. However, as the authors themselves point out, a student can be in several Quadrants simultaneously, and evaluating what kind of support would be beneficial for their learning at that specific moment remains a very complex endeavour.

2.2.2.2 D'Mello and Graesser Zone of Optimal Confusion

D'Mello and Graesser have been interested in emotional processes generated by the cognitive incongruity and how cognition and emotion interact during learning. They define *confusion* as the 'emotional signature' of cognitive imbalance: it is a central aspect of complex learning.

Current theories postulate that confusion emerges through a mismatch between incoming information and existing knowledge while processing new information. Information that can be assimilated or accommodated without too much restructuring does not induce a state of confusion, whereas more significant discrepancies trigger a relevance alarm during the appraisal process, mobilising attentional resources. The resulting emotion may differ depending on the context: at negative valence and moderate arousal, it appears to be confusion.

Whether it is the work of Gross (2008) on confusion regulation strategies, Piaget's description of maladaptive or adaptive responses to cognitive conflict, or Festinger's (1957) cognitive dissonance, it seems to be accepted that confusion is perceived as a state that requires cognitive action to resolve, either through its negation or through adaptive strategies. This claim is of interest in educational contexts: several studies, including D'Mello, Lehman and Graesser (2013), have shown that the resolution of even partial confusion is essential for learning. For example, Van Lehn et al. (2003) report that the quality of learning outcomes is strongly influenced by the presence of deadlock in the process. Craig et al. (2004) even showed that confusion is the emotion most positively correlated with learning: "The effect size on learning (0.64) observed when confusion was present versus absent suggests that some level of confusion is critical for optimal learning".

However, the benefit of confusion for learning depends on its efficient regulation: confusion in itself signals to the organism that there is something relevant to assess, which has the effect of mobilising attentional, cognitive, and motivational resources to process the information. However, it is not the occurrence of confusion, but the choice of a strategy to cope with confusion that will determine whether learning takes place or not. The motivation to reduce dissonance will lead individuals to problem-solving (alone or with the group) to return to a state of equilibrium. Therefore, the cognitive activities related to the state of confusion, not confusion itself, seem to have a tangible impact on learning. Confusion plays a role as a moderating variable - this comment is valid for a state of confusion that leads to further questioning and not confusion that leads to frustration.

From this research, the authors deduced that situations where confusion is productive are during complex learning or from a certain level of expertise. On the other hand, persistent confusion can be detrimental and lead to failure for learners who are already struggling or demotivated, in situations with high risks of dropping out (D’Mello & Graesser, 2012). They thus hypothesise a *Zone of Optimal Confusion*, where learners are confronted with content that is sufficiently complex to interest them - in line with Paul Silvia's work on interest (Silvia, 2008) - yet within their reach, to avoid slipping into frustration or boredom. Figure 3 illustrates this Zone of Optimal Confusion in relation with other emotions.

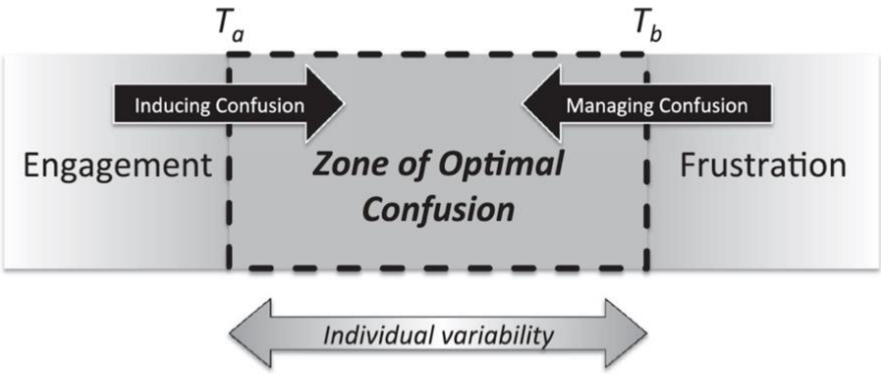


Figure 3. A representation of the Zone of Optimal Confusion by Arguel and Lane (2015)

Therefore, D'Mello and Graesser (2013) propose to adapt the level of confusion to the learner's specific needs. That type of individualisation of learning can be extremely difficult to implement. Thus, the authors argue that these needs can be met through computer developments, with the production of systems that elicit confusion and provide the appropriate levels of support to keep learners in their optimal zone of confusion.

As in other studies, the key idea here is that confusion is only beneficial when productive, not persistent. This outcome relates strongly with Productive Failure as an instructional design model that takes disruption and its regulation into account.

We have just seen that confusion is most appropriate and stimulates learning in complex learning or for learners with a certain level of expertise, but that it can, on the other hand, be deleterious and lead to failure for learners who are already struggling. We can thus question the notion of failure, as it is still widely perceived today. The next theory addresses these issues.

2.2.2.3 Productive Failure

Productive Failure (PF) is an educational design model first proposed by Kapur and Bielaczyc (2012). It is rooted in a larger method called PS-I, i.e., *Problem-Solving followed by Instruction*

(Loibl et al., 2017), as a way to foster engagement and support learning, as opposed to Direct Instruction (DI) which is still the most dominant method in schools and universities.

PF posits that the Problem-Solving phase should be designed to intentionally lead to suboptimal or inaccurate solutions, thus preparing the mind for the upcoming learning from instruction. It relates to Kort et al. criticism of overpolished classes that hinder failure and confusion, thus preventing the appropriate emotions to emerge during the learning cycle, further hindering cognitive regulation that leads to learning. Their research has shown that the more incorrect solutions the students produce, the better they are at conceptual understanding and transfer of the acquired knowledge.

Several studies (Kapur, M, 2016; Sinha & Kapur, 2021) have addressed the contexts in which Productive Failure works from the ones where it fails. Some of these dimensions are consistent with other work on confusion. One of the main interests with Productive Failure is that it offers a framework to design course activities that can include student's self-report on their confusion level, which would then allow the teacher or system to provide an adapted response to their exact needs.

Kapur (2016) has identified a set of factors that influence the positive use of PF in instructional design. Among these, the ability of one to accept doubt and failure seems of particular interest for our study. Risk-aversion, tolerance to errors, and a global culture that fosters trying, failing, and learning from mistakes are major influences on the quality of learning. The next chapter gives an overview of the main theories on these topics.

2.2.3 Motivation, mindsets, and risk-aversion

2.2.3.1 Errors as a mean to improve

Presented by Carol Dweck in 2006, the mindset theory seeks to understand why some students fail when others strive, particularly in the face of difficulties and challenges during learning. To address this question, they set out to analyse several factors linked with academic performance, such as goals, attributions, and motivation. While working on this, they realised that all these variables could be organised in a single system, namely "meaning systems" that are strongly influenced by mindsets (Molden & Dweck, 2006). For Dweck and colleagues, mindsets are beliefs about one's abilities. In an academic context, a *growth mindset* means learners believe intelligence can be developed, as opposed to a *fixed mindset* in which learners believe intelligence cannot evolve, it is *fixed*. Research on these mindsets have consistently shown that learners with a growth mindset outperform learners with a fixed mindset (Dweck & Yeager, 2019).

These findings are thought-provoking and of high interest for our research: if confusion is triggered by an imbalance and breaks the flow of understanding, people with fixed mindset could be very sensitive to the impression of failure and may want to avoid this situation to maintain and protect their self-esteem. Intuitively, they would not strive in a confusion-based teaching approach. On the other hand, learners with a growth mindset might strive in this situation as they welcome doubt and errors as a way to improve their abilities. The question therefore is: do mindsets and their elements have an influence on the way learners react to confusion?

We can dig even deeper in these meaning systems: a crucial sub-element for our research is *effort belief*: the belief that effort is actually a positive process to help grow an ability as opposed to the demonstration of an inability that cannot be improved. In that sense, as confusion generates the need to produce effort, a fixed mindset would hinder the call to the proper cognitive resources essential for learning. As Dweck and Yeager (2019, p. 483) state, for learners with a fixed mindset, "high effort may more readily be seen as indicating low ability, and setbacks are more easily attributed to low ability. When this happens, persistence can be curtailed". Whereas people with a growth mindset will perceive setbacks as information, constructive and vital feedback on their learning journey.

Setting mindsets in parallel to Productive Failure, we see an emerging pattern in all these studies: most of the outcomes are predicted by one's ability to withstand errors, an ability we may call *tolerance to error*. Substantial research has been conducted, in particular in the field of social psychology on risk-aversion. As we stated, learning means taking a risk, people who are risk-averse may choose to delay learning and feeling confused only until they have no other choice. This relates to the effort belief previously stated: as effort is perceived as negative and showing inability, they will try as much as possible to avoid situations that require effort.

To measure this *tolerance to error*, we refer to the work of Lauzier on Error Orientation (Lauzier, 2011). Lauzier's work has revolved around the effect of the perception of error in corporate employees on their motivation to learn and intentions to transfer knowledge in their daily job. His research is of particular interest for our study as error orientation is a strong predictor of knowledge transfer, thus fostering deeper learning. Lauzier and Mercier (2018) found that motivation has a mediating role between learning by error and the intention of transferring learning. As such, we can infer that a learner with a positive error orientation may welcome confusion as an indicator of development and may react more positively to it, thus fostering better learning outcomes.

We translated and used a subset of items on the Error Orientation Questionnaire (Rybowiak, Garst, Frese & Batinic, 1999) used by Lauzier and Mercier in their study (Lauzier & Mercier, 2018). Details are further described in the method section.

2.2.3.2 What we had to leave out

In this research, we focus on the emergence of confusion in the learning phase. The complexity of the process has led us to consider focusing on two specific aspects, the confusion and the perception and ability to accept doubt and failure.

Therefore, we had to leave out interesting elements like the type of motivation (intrinsic vs extrinsic) for the topic being studied. Once again, the link with the appraisal phase of relevance triggers motivation. Intrinsic motivation positively impacts learning and affective state is influenced by level of motivation. Even though this research does not focus on motivation, we added three items to measure interest and motivation to the survey to control for a possible complete lack of interest in the topic presented, that would influence the experience.

Akin to motivation theories, we also left out Deci and Ryan's Self Determination Theory (SDT), although we hypothesise that this theory might correlate with learner's perception of error and its use to learning, as feelings of competence and autonomy grow.

Finally, Flow theory (Csikszentmihalyi, 1990) is very popular, and it would be interesting to understand if and how it is linked to confusion. For instance, using D'Mello and Graesser's model of Zone of Optimal Confusion, we could legitimately ask whether it is experienced as flow by learners. This is left aside for further studies, as discussed in the last chapter of this paper.

2.3 Digital-based Learning

To conclude this introduction, we must talk about the research in instructional design in digital environments. The scope of this research is on autonomous learning in an online course, such as an e-learning module or a MOOC. There is an important body of research on social influences and collaborative learning. It can be debated to what extent giving feedback is a social dimension or not. In our study, we will deliberately exclude any social factors, to concentrate on the learner's individual experience with the content. Interesting use of artificial intelligence and adaptive learning in online environments will be addressed in the discussion section.

Digital learning is highly influenced by studies of human-computer interactions. Since the generalisation and access to internet has shifted part of the classical classroom instruction to computer-based online training, the importance of the learner's experience with the system strongly impacts the emotions they feel when they learn.

Richard Mayer's multimedia principles designs (e.g., Mayer, 2009: Mayer, 2017) are widely used today as a methodology to design learner-centred courses as opposed to technology-centred classes where trainers tend to forget learning outcomes in favour of technological gimmicks. We have implemented main principles in our design to minimise the risk of a bad instructional design influence on the emergence of emotions.

After primarily focusing on cognitive function, like memory and attention, researchers got interested in technologies that can effectively detect and respond to users' emotions. Affective Computing (Picard, 1997) has paved the way for studies on such systems in the education field, in particular Affect-Aware learning technologies and Intelligent Tutoring System. Intelligent Tutoring Systems are educational applications that are used to automate and adapt the scaffolding a teacher would do to the learner's needs. They have been used successfully in research on confusion to demonstrate a positive correlation between learning and confusion (Craig et al., 2004). In this work, we did not develop ITS due to technological constraints, but we use outcomes from the research to strengthen our instructional design, namely we created dialogues as used in AutoTutor (D'Mello & Graesser, 2013).

2.3.1 Induction of confusion in e-learning

Research about confusion has explored an important number of methods to induce, detect and measure, as well as help regulate confusion. The choice to actively induce confusion depends on two important elements: On the one hand, well-regulated confusion can be a key, yet not necessary, part of learning. Moreover, as already discussed, confusion might not be appropriate for a certain profile of students or certain type of instruction. On the other hand, not trying to design for confusion may mean that it could never occur, thus limiting the learning outcomes.

In this work, we based ourselves on two studies that use dialogues and misconception to design with confusion (D'Mello & Graesser, 2013; Muller, 2008). We present below a brief overview of the current state of the research that led to our choice of design.

Several methods have been used to elicit the feeling of confusion in experiment designs (for a review, see Sullins & Denton, 2019).

- **Anomalous information**: Limon (2001) showed that the presentation of anomalous information was helpful not only to experts, but also to learners with a low-domain knowledge. They conclude that confusion triggers thought processes that "serve as a catalyst in the process of conceptual change" (Sullins & Denton, 2019, p.291).

- **Contradictory information**: Research in Intelligent Tutoring System often used contradictory information to induce confusion (Lehman et al., 2012; D'Mello & Graesser, 2013), with scenarios using two pedagogical agents disagreeing during a dialogue. Results showed that contradictions managed to trigger confusion, only if the contradictions were sufficiently severe.

- **Breakdown**: A third type of confusion-based design used in STEM (D'Mello & Graesser, 2014b) is to present learners with a diagram of an industrial process or machine for a couple of minutes. The next phase contains the presentation of the same diagram with a problem during the process ("breakdown"), thus inducing confusion and deeper thinking. This method has proved to have better learning outcomes, also in the long term, compared to regular direct instruction. These findings are supported by impasse-driven learning theories (VanLehn et al., 2003).

Our choice for this work is detailed in the Method section.

2.3.2 Synthesis of confusion and impacting factors

As we saw in this chapter, confusion can be considered an epistemic emotion, namely an emotion that arises during learning, deeply linked with cognitive states and that its occurrence can be linked to an imbalance between new information and existing mental models. Being able to detect confusion and provide appropriate support seem to predict deeper learning and stronger retention.

Designing systems like ITS seems like a promising idea to induce and help regulate confusion, but too costly at this stage for most instructional designers. Moreover, tools to detect the occurrence of confusion like IA-based apps that read actions units are interesting for research and getting better every year. However, it is unlikely that they will be available to teachers and trainers any time soon.

Our approach for this work is therefore to use the models and principles described earlier to define a confusion-based instructional design that produces course scenarios fostering the emergence of confusion.

3. Focus of the research

After reviewing actual research in the field of affect in education, there is great temptation to include lots of factors that seem to positively influence high-quality learning. Those countless influences, stemming from cognitive, affective, or social perspectives are undoubtedly important, but would generate too much complexity in the design.

We thus decided to focus on following elements:

- **Individual learning**: although confusion could be induced through social activities such as debate to generate socio-cognitive conflict, we decide to focus on cognitive conflict within a learner as an autonomous agent. As more people learn with online resources, the resort to peers or tutors is less likely (e.g., in MOOCs with thousands of unknown and unrelatable enrolled users).
- **E-learning module**: for the same reason, we decided to study emotions in a stand-alone e-learning module, although current trends on Blended Learning designs may favour the induction of confusion during specific phases of the instruction.
- We decided to design for **confusion**, in its acceptance as *the emotional signature of cognitive disequilibrium* and a valid precursor to deep learning (D'Mello & Graesser, 2012).
- Given the situation (Covid pandemic), we resort to **self-reports on emotions**, as it is more straight-forward to implement and less invasive than other methods described above, like filming learners and/or ask observers to code the emotions.

3.1 Research question and hypotheses

As the previous section showed, discarding counter-intuitive approaches with negative emotions may hinder the design of better instruction, as it would leave out confusion. This is especially true in e-learning designs as these are often very linear, and it is more difficult to provide adaptive support. It seems therefore important to study if and how confusion-based instructional design for e-learning modules can be beneficial to learning, and in which contexts.

Because the body of research from these past ten to twenty years seem to reinforce that confusion is beneficial to learning, the next step would be to understand in what way, and what other variables are impacted that could be directly linked to a better learning experience.

Our general research question is thus: How can a confusion-based instructional design foster learning? To explore this question, we first try to replicate findings regarding productive confusion (D'Mello & Graesser, 2012). Then, we add several parameters in interaction with the confusion-based design to test the impact of error orientation styles to identify in what conditions this design is beneficial for learning. To our knowledge, no studies have tried to link the reaction to confusion with error orientation styles. Research on Productive Failure (Kapur, 2016) has mentioned the importance of mindsets, but no direct link has been made with confusion.

Our main research question is as follows: Does a confusion-based instructional design promote better learning than classic direct instruction (D'Mello & Graesser, 2012; Kapur, 2016)? Three hypotheses arise from this question, that we will test in the experiment:

1. **The feeling of confusion triggers cognitive processes that are beneficial to learning** (D'Mello & Graesser, 2012): As discussed, it is not confusion as such that drives learning. Confusion triggers the activation of cognitive resources like attention and working memory for a deeper processing of the information. As such, learners reporting confusion should display better learning outcomes than those who do not report having been confused.

2. **The feeling of confusion fosters learners' confidence**: we want to explore if a confusion-based design would promote confidence in one's own knowledge, as confused learners would supposedly invest more work in the acquisition of the knowledge, thus feeling more confident in their new mental models. To test this hypothesis, we measure the confidence reported by learners before and after the instruction.
3. **Confusion-based instructional design is only beneficial if learners have a positive error orientation style:** As seen with Dweck's Mindsets (Dweck, 2006) or Error Orientation Styles (Lauzier & Mercier, 2018), learners may respond to confusion in conflicting ways depending on their perception of error and failure in the learning process. As confusion creates an imbalance, learners with a negative error orientation style may feel more anxious than they would feel confused, thus not learning as well as the learners with a positive error orientation style.

The next section on method will describe how we have operationalised these hypotheses and give details on the methods and theories used to create a confusion-based instructional design for our study.

4. Method

4.1 Participants

Primary target population are university students, from the universities of Geneva, Lausanne and UniDistance. They received the link to the course either by email or by scanning a QR code after a quick presentation of the context. The content of the course, First Aid, was chosen in order to control for motivation and prior knowledge of this extended population (discussed below).

This work has been approved by the Ethics Committee of the University of Geneva. Participants signed an agreement to participate in the study and were allowed to leave at any time during the study. All data was collected anonymously. They did not receive any credit or other reward for their participation.

Fifty-four participants took part in the study: 38 of them identify as women, 10 as men, 1 as other and 1 chooses not to say. Average age of participants is M = 34.00 (SD = 10.96) for women, M = 34.78 (SD = 8.67) for men. Neither gender, nor age was controlled for group assignment as we have deemed them irrelevant for this experiment.

Four participants did not finish the course and were thus removed from the sample, with a final N = 50. We made a preliminary calculation to determine the sample size for significant results. Using G*power, we determined that the sample size should be 126, which was unlikely due the lack of incentives (no credit or money) and the difficulty to get in touch with people during the pandemic. We decided to stop the collection four months after the start of the recruitment.

4.2 Material

4.2.1 Data Collection

We used the survey application Qualtrics for all questionnaires, as well as the course. We considered using authoring tools or Learning Management System like Moodle to create a more realistic training experience, but this did not allow us to track all variables in a unique environment. We thus chose Qualtrics as the sole platform to minimise the risks of external factors interfering with the experience.

4.2.2 Course Design

To explore the impact of confusion on students, we designed an experiment contrasting educational content designed to induce confusion (CC) and a more classical direct instruction approach, not designed to induce confusion (CSC). It should be noted here that we can only control the induction of confusion in the conception of the course scenario. We cannot control the occurrence of confusion in learners, as it could arise even with regular instruction design, if the information is relevant enough and incongruent to their models. For this reason, we use a self-report questionnaire to measure the occurrence of main epistemic emotions during the course.

As Muller (2008, p.18) states in his thesis, “many researchers claim it is impossible and unproductive to attempt controlled experiments in education since the number of variables that may impact on learning in authentic education settings is so great”. Although design-based research approaches are getting a stronger focus in research in education settings, in an attempt to include the complexity of all the variables that can affect learning, we decided to design an experience in order to have, as much as possible, a clear cut-off between the confusion setting and direct instruction.

We have developed a design aimed to balance the trade-off of an authentic approach with a better controlled environment to detect emotional response and level of confidence during the learning process.

4.2.2.1 Induction of confusion

As presented in section 2, several designs have been explored in research on confusion induction. In this study, we based ourselves on two studies that use dialogues and misconception to design with confusion (D’Mello & Graesser, 2013; Muller, 2008).

We decided to avoid using contradictory information from two pedagogical agents, as we thought this could affect the confidence in the course. Therefore, we opted for a design based on anomalous information, in the form of misconceptions between two neutral characters.

Some authors propose to use natural occurrences of confusion, known to be frequent in complex learning, and give an adaptive response, based on correctness and certainty (Forbes-Riley & Litman, 2011). Due to the objectives of this work, we could not risk missing the trigger of confusion. We thus designed for confusion but kept the measurement of a confidence level in order to detect changes in confidence based on confusion.

4.2.2.2 Detection and measurement of confusion

Real-time affect detection (Praiva, Prada & Picard, 2007) is a popular research topic, but due to technological constraints, research still relies heavily on self-reports to detect learners' emotions. This implies that participants can detect and name their emotions as they occur, and that the meaning of *confusion* or *interest* is similar throughout the sample. As usual in these experiments, biased perception of external factors is to be considered when discussing results.

We left out computer detection and AI-based detection (Bahreini, Nadolski & Westera, 2014; Bosch, Chen, & D'Mello, 2014; Ismail & Syaiful, 2015). Although it would be highly interesting to remove bias from self-reports of learners in automating the process, the process is still too invasive to feel transparent to learners and may significantly impact their emotional states.

Finally, other researchers trained coders to observe participants and code their emotions, either alone (Craig et al., 2004) or in comparison to learners' self-assessments of a filmed study session (Tiam-Lee & Sumi, 2019). This method could ensure a stronger inter-rater reliability on induction, but it is costly to train coders and it still relies on human interpretation.

In this experiment, we used a classical self-report on five main epistemic emotions (interest, surprise, confusion, frustration, and boredom) with a 5-point Likert Scale to detect if confusion was triggered and if other emotions were present.

4.2.2.3 Course Creation

As this was not a real-life setting and we had to recruit participants who had no known prior interest for the training, we decided to create a short course to keep motivation, attention, and interest to participate in the study as high as possible. This was crucial to ensure the induction of confusion during learning. Without any motivation, confusion is unlikely to appear. With this constraint in mind, we made a certain number of choices to foster the chances of confusion in our study.

We decided thus to systematically control for major aspects in Instructional Design, namely:

1. Presentation of content following Richard Mayer's principles of multimedia for e-learning design (Mayer, 2017).
2. Motivation and focused attention through use of main gamification principles for education (Kapp, 2012).
3. Interest and engagement by choosing a training content that anyone could relate to and allow for semi-authentic situations.

4.2.2.3.1 Multimedia Design

Richard Mayer's principles for multimedia are one of the leading frameworks in instructional design today. We used an updated list specifically aimed at using multimedia for computer-based environments (Mayer, 2017). Thus, the training material in the experiment has been created with following principles:

- **Coherence**: the situations were kept very simple, both in text and images. Only relevant information was included to reduce extraneous cognitive load and keep attentional resources focused on the information to be learnt.
- **Signalling**: no signalling was used to avoid creating differences between the confusion and no confusion conditions. However, due to the design as comic strips, the texts were kept simple and short, thus rendering signalling of information less critical.
- **Segmentation**: presenting the content in short chunks of information, to reduce cognitive load and help learners focus on the relevant information to be learnt.
- **Pre-training**: the initial diagnostic quiz serves multiple purposes: not only does it allows us to determine the level of prior knowledge, but it also helps learners to activate the right kind of mental models to prepare for the subsequent training.

At this point, a valid critique could be made that the design does not include images that describe the training material, although this would be consistent with Mayer's multimedia principles. The reason for excluding images related to the content was to control for the confusion effect. Although in a real-world design, we might have wanted to use images to support the text, in this experiment, it was assessed as a risk for the induction of confusion. Thus, we chose to keep design to its simplest form, pondering that the short length of the micro-learning would compensate for the negative impact of not using more multimedia assets.

Principles not mentioned above (spatial contiguity, temporal contiguity, redundancy, personalisation, voice) were discarded as irrelevant to the present design, mostly due to the short length and very simple design of this micro-learning.

4.2.2.3.2 Motivation and attention control via gamification principles

Using Kapp (2012) game-based methods for education, we reviewed our training and ensure learners engagement by:

- Using stories instead of a more traditional lecture-type format, to render the material as close to real-life as possible, thus allowing immersion, known to promote deeper understanding and engagement. The dialogues between agents mimic interactions that learners could encounter in their daily life.
- Choosing a theme that would appeal to most participants (First Aid), with the goal to let intrinsic motivation for the content overcome the extrinsic motivation of just participating in a study for the sake of it. This also allows control of different levels of prior knowledge: complete beginners would discover best practices, while intermediate or expert learners could check the accuracy of their current knowledge.

4.2.2.4 Course content

Based on Muller (2008), we decided to use *misconceptions* as a “simple” way to induce confusion. Misconceptions are defined as a false belief or an opinion that is incorrect due to faulty thinking or lack of understanding. In line with research on confusion, it is likely that if learners had a misconception, they would be confused if the content of the course was incongruent with their current knowledge. Muller (2008) used this approach to study learning outcomes in physics education. His results show that if learners are given classical direct instruction, they do not change their misconceptions, whereas if they are shown dialogues in which one of the protagonists explicitly states the misconception which is then questioned and corrected, they have a significant increase in learning outcomes (Muller, 2008). Other research in Affective Computing and Intelligent Tutoring System (D'Mello & Graesser, 2013) have used dialogues and discrepancies during dialogues to induce confusion.

To foster relevance and ensure existing misconceptions in a varied sample of participants, we decided to use First Aid as a topic for the course. First Aid can be of interest for any human being. We framed the recruitment text with the aim to appeal to both beginners and experts, in emphasising the game-like approach to test one's own knowledge on First Aid, that could one day save a life. This topic was also selected as it is highly prone to misconceptions. Most people are learning about First Aid with their families, thus having a strong common sense of what should be done, because "my grandmother told me when I was little". The misconceptions in First Aid were selected through an internet search, followed by a validation by a physician.

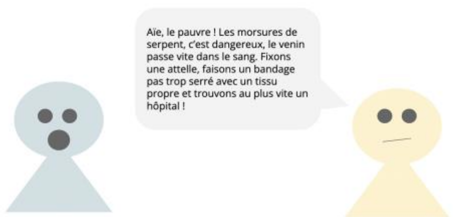
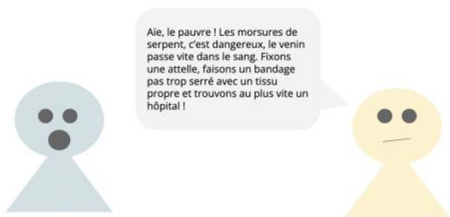
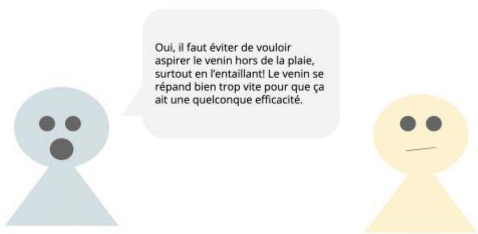
4.2.2.4.1 Scenario

We created eight dialogues for eight topics based on the most widespread misconceptions in First Aid. These topics are (1) Choking, (2) Nosebleed, (3) Burns, (4) Seizure, (5) Snake bite, (6) Poisoning, (7) Fever and (8) Cardiac arrest.

All dialogues are designed in the same format, i.e., four images containing two characters discussing a real-life incident. Each image represents one character speaking in turn. We alternate the character that starts the dialogue between each topic to avoid risks of detecting a pattern.

Dialogue steps	Confusion-based condition	Control condition
1	Description of the incident	
2	Correct solution	
3	Misconception: disagreement with the previous statement	Agreement on correct solution, with introduction of misconception as a wrong approach
4	Repetition of correct solution: Explanation as to why misconceptions is wrong	Repetition of correct solution with agreement on previous explanation

To ensure an equivalent processing of the content in both conditions, we used the same design and balanced the number of words to avoid differences in cognitive load or amount of information received. Example for snakebite is shown in Figure 4. Other examples of the dialogues can be found in Appendix A.

	Confusion-based Dialogue (CC)	Direct Instruction Dialogue (CSC)
1		
2		
3		

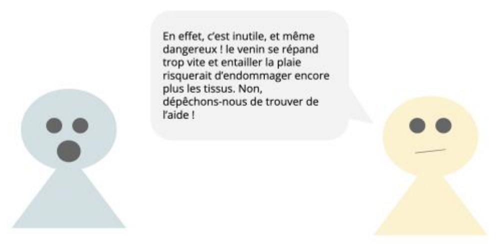
4		

Figure 4. Course content for topic 5: Snakebite

4.2.2.5 Questionnaires and scales

Error Orientation Scales

The study includes a short socio-demographic questionnaire (age and gender) followed by a subset of 8 items taken from the Error Orientation Questionnaire (Rybowiak et al., 1999). They have been adapted to the course context and translated into French. Items 1 to 5 describe the emotions felt towards errors, such as "I find it stressful when I err" or "I feel embarrassed when I make an error". Items 6 to 8 measure the perception of error as a means of learning. They include questions such as "My mistakes help me to improve my learning". All items are grouped into the single variable "Error Orientation".

Attitude Survey

The questionnaire to test interest for the topics is adapted from the Attitude Survey (Huang, 2017). We used a subset of 3 items, with a Likert scale from 1 to 5 (1 = "I strongly disagree" to 5 = "I strongly agree"). Items include sentences such as: "I enjoyed learning more about First Aid" and "I find that what I learnt is this course on First Aid could be useful".

Finally, the mental effort was measured with the Cognitive Load Scale (Paas et al., 1994), containing a single question: "I had to invest a mental effort…", using a 9-point Likert scale, ranging from 1 = "very, very low" to 9 = "very, very high".

Questionnaires and scales can be found in Appendix C.

4.3 Experimental procedure

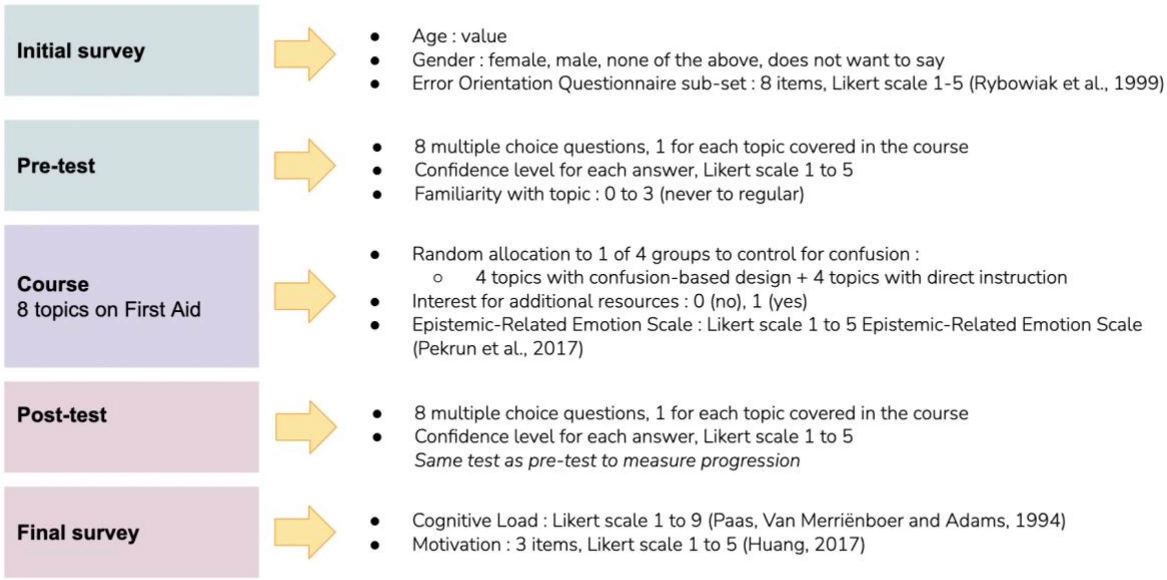


Figure 5. Experimental procedure synthesis with phases, content, and scales.

Participants were directed to Qualtrics with a URL or a QR Code. They were encouraged, though not compelled, to do the study on a computer or a tablet, as images were small on smartphones. They could do the study any time, on their own. The procedure, as summarised in Figure 5, took a total average time of 20 minutes distributed as follows:

1. **Initial survey**: age, gender and 8 items of the Error Orientation Questionnaire from Rybowiak et al. (1999). The Error Orientation Questionnaire contains 8 scales used to measure how people react and cope with errors. Using an adaptation to education from Lauzier & Mercier (2018), we decided to focus our research on two factors: *error strain* and *learning from errors*, including the 8 items selected for our work.

2. **Pre-test**: 8 questions for each topic on the course on First Aid, as a multiple-choice question with 4 possible answers: the correct answer, a distractor close to the correct answer, the common misconception on the topic, a distractor close to the misconception. Below is an example for topic 5 “snakebite”:

“ Vous admirez une vue magnifique depuis un surplomb rocheux avec une amie, lorsque celle-ci se met à crier en se tenant la main. Vous voyez déguerpir un serpent apeuré, mais n’avez pas le temps de bien le voir. Votre amie s’accroupit, visiblement choquée. Que faites-vous ? ”

- Je procède à un bandage léger, fixe le membre avec une attelle si possible et je trouve au plus vite un moyen de rejoindre un hôpital *(correct answer)*
- Je fais un garrot bien serré pour empêcher le venin de propager et je trouve au plus vite un moyen de rejoindre un hôpital *(correct answer distractor)*
- Je pratique une petite entaille au niveau de la morsure, aspire le venin pour éviter sa propagation rapide, et je trouve au plus vite un moyen de rejoindre un hôpital *(misconception)*
- Je pratique une entaille pour aspirer le venin, poursuis par un garrot bien serré, et je trouve au plus vite un moyen de rejoindre un hôpital *(misconception distractor)*

For each question, we measured a level of confidence (1 = “not at all confident” to 5 = “totally confident”) and the degree of familiarity (0 = “I have never been in this situation” to 3 = “I have regularly experienced this situation”) to determine prior knowledge. Other examples of questions are listed in Appendix B.

3. **Course**: To control for pattern detection, participants were randomly assigned to 4 groups with different instructional designs (see Table 1 below). Each dialogue was created in two versions: with confusion-based design (CC) and in a regular direct instruction design (CSC). Participants were randomly seeing 4 topics in CC and 4 topics in CSC. After each topic, we asked participants to self-report on 5 epistemic emotions (interest, surprise, confusion, frustration, and boredom), with a Likert scale (1 = “not at all” to 5 = “completely”) as well as on their interest in getting additional resources on the topic (1 = “yes”, 0 = “no”).

4. **Post-test**: The post test is the exact replica of the pre-test, except for the exclusion of the questions on familiarity, that are not relevant to compare.

5. **Final survey**: After the post-test, participants rate their interest for the topic on 3 items (Huang, 2017), as well as rate their global mental effort based on the 9-point Likert Cognitive Load Scale in instructional research (Paas, Van Merriënboer & Adam, 1994).

If the learners had replied “yes” at least once for additional resources, a page containing links to documents and videos for the chapter was displayed before thanking and taking leave of participants. We chose not to give the additional resources immediately after each topic to avoid differences in time spent on course and cognitive load between participants.

No feedback was given on the pre-test in order to avoid participants learning from feedback instead of the course content. The answers to the post-test and thus the learner’s progression is due solely to the course content.

To evaluate the experiment, we asked 5 students to do a pre-testing of the study. This test validated the choice of misconceptions on all 8 topics. Qualitative feedback contained hints of confusion on the content: “All these years, my parents and teachers have been lying to me ?!”, but also on the form: “I was not sure what the right answer was”. We kept the test as designed initially.

4.4 Experimental design

To control the order effect, we use the counterbalancing method: all participants went through 4 topics with Confusion-based Design (CC) and 4 in Direct Instruction (CSC). They were randomly assigned into 4 groups following rules shown in Table 1.

Table 1. Group randomisation in experimental condition (CC vs CSC)

Group	**Topics with Confusion-based design (CC)**	**Topics with Direct Instruction (CSC)**
1	1, 4, 5, 7	2, 3, 6, 8
2	2, 3, 6, 8	1, 4, 5, 7
3	3, 5, 6, 8	1, 2, 4 ,7
4	1, 2, 4 ,7	3, 5, 6, 8

4.4.1 Variables

4.4.1.1 Independent variables

To measure the effect of a confusion-based design and its relation to error orientation style, we use two independent variables:

Design Type (VI1): A within-subject variable with two modalities: Confusion-based (CC) and Direct Instruction (CSC).

Error orientation (VI2): a between-subject variable calculated on a score (mean) on 8 items with a 5-point Likert Scale, ranging from 1 = “Totally disagree” to 5 = “Totally agree”. It has two modalities: Low tolerance to errors (LT), including scores from 1 - 2.999, and High tolerance to errors (HT) including scores from 3 to 5.

There is an ongoing debate on how to measure central tendency on a 5-point Likert scale. Some researchers point out that a mean between “totally agree” and “agree” does not mean anything, whereas other researchers state that the median is likely to span a wide range of actual percentile values, losing granularity to show differences between groups, as thus, argue to use the means. A comparative study of t-test versus Mann-Whitney-Wilcoxon (de Winter & Dodou, 2010) has shown that there are no significant difference or risk of Type 1 errors between the two tests and that t-test can be used confidently. Therefore, we decided to compute items using the mean in order to minimise the loss of information in the process of grouping.

4.4.1.2 Dependent variables

Learning gains: a score calculated by subtracting the score on the pre-test from the score on the post-test, ranging from -4 (all correct answers on pre-test and none on post-test) to + 4 (all correct answers on post-test and none on pre-test).

Confidence Level: a score (mean) on a 5-point Likert scale, ranging from 1 = “not confident at all” to 5 = “Totally confident”.

Emotions: a score (mean) for each of the 5 epistemic emotions chosen for this study, namely Interest, Surprise, Confusion, Frustration and Boredom. Based on a 5-point Likert scale, ranging from 1 = "totally disagree" to 5 = "Totally agree" as an answer to the question "During this course, I felt *emotion*…".

4.4.1.3 Other variables measured to control induction of confusion

Prior knowledge: a score on each topic, ranging from 0 = "I have never experienced this situation" to 3 = "I have experienced this situation on a regular basis".
As confusion arises when an incongruity is detected, it requires the existence of prior models for comparison and generating a discrepancy. If confusion is felt by beginners, it may imply that the information is too complex to be processed with their current knowledge and does not trigger productive confusion.

Interest: a computed variable (mean) on 3 items with a 5-point Likert-Scale, ranging from 1 = "not interested at all" to 5 = "highly interested". If learners are not interested or motivated by a topic, it is highly unlikely that they would feel confusion as it would not pass the threshold of relevance.

Additional resources: a computed variable (mean) on the question: "Would you like to know more about this topic", with 1 = "yes" and 0 = "no". This variable serves as an indicator of interest, that could contribute to engagement and mobilisation of cognitive resources.

Mental Effort: a score on the question "I had to invest a mental effort…", using a 9-point Likert scale, ranging from 1 = "very, very low" to 9 = "very, very high". This variable allows us to analyse if people reporting being confused more frequently have a higher mental effort score, as would be expected, since confusion triggers the activation of cognitive resources.

4.5 Statistical Analysis

We used Jamovi and RStudio to analyse our data, after a phase of clean-up and recoding in Excel. Statistical tests used are indicated in the results in each section.

5. Results

5.1 Descriptive analysis

5.1.1 Participants

Sample characteristics like gender and age are already described in the method section. As discussed in the same section, we measured two variables influencing the success of the induction of confusion:

- Interest for the topic: If prior interest is low, relevance will not be triggered, and confusion is unlikely to appear. Participants reported a high score (M = 4.37; SD = 0.65) on a 5-point Likert Scale, with 5 being highly interested.

- Prior knowledge: we measured prior knowledge as a degree of familiarity with the First Aid topics presented in the course. Participants reported a low level of familiarity with the topics (M = 0.35; SD = 0.27), ranging from 0 = "never experienced" to 3 = "experienced on a regular basis". There are small non-significant variations between topics. Participants are therefore considered as novices.

5.1.2 Dependent Variables

We calculated the standard deviation and the mean for each of the dependent variables (Learning gains, Confidence Level and Emotions), according to the Design Types, shown in Table 2. We made the same calculations on dependant variables for Error Orientation (Table 3) and the interaction of both factors (Table 4).

Table 2. Means, Standard Deviations for Learning Gains, Confidence Level and Emotions, split by Design Type (Confusion-based (CC) - Direct Instruction (CSC))

	CC		CSC	
	M	*SD*	*M*	*SD*
Learning gains	1.280	1.262	1.240	1.153
Confidence Level	0.935	0.721	1.160	0.717
Interested	4.180	0.786	4.198	0.774
Confused	2.020	0.954	1.755	0.889
Surprised	2.470	0.914	2.125	1.055
Frustrated	1.503	0.824	1.370	0.753
Bored	1.415	0.779	1.370	0.761

Table 3. Means, Standard Deviations for Learning Gains, Confidence Level and Emotions, split by Error Orientation (High Error Tolerance (HT), Low Error Tolerance (LT))

	HT		LT	
	M	*SD*	*M*	*SD*
Learning gains	1.364	1.432	1.231	1.139
Confidence Level	0.989	0.730	1.064	0.727
Interested	4.080	0.729	4.220	0.790
Confused	1.932	1.036	1.875	0.901
Surprised	2.261	1.076	2.308	0.981
Frustrated	1.477	0.879	1.425	0.766
Bored	1.648	0.938	1.321	0.701

Table 4. Means, Standard Deviations for Learning Gains, Confidence Level and Emotions, split by Design Type and Error Orientation.

	CC				CSC			
	HT		LT		HT		LT	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Learning gains	1.545	1.440	1.205	1.218	1.182	1.471	1.256	1.069
Confidence Level	1.023	0.958	0.910	0.653	0.955	0.445	1.218	0.772
Interested	4.045	0.705	4.218	0.811	4.114	0.786	4.222	0.779
Confused	2.091	1.091	2.000	0.927	1.773	1.003	1.750	0.868
Surprised	2.455	1.089	2.474	0.875	2.068	1.079	2.141	1.062
Frustrated	1.523	0.945	1.498	0.800	1.432	0.852	1.353	0.734
Bored	1.705	0.993	1.333	0.701	1.591	0.924	1.308	0.710

5.2 Effects on Learning gains

5.2.1 Main effect of Design Type

As can be seen in Table 2, participants showed consistent Learning gains on both Design Types. On a range from 0 (= no correct answer) to 4 (all correct answers), scores increased on an average by 1.28 (*SD* = 1.26) in the CC condition, from *M* = 1.88 in the Pre-test to *M* = 3.16 in the Post-test. On the CSC condition, we found similar results, from *M* = 2.02 in the pre-test, to *M* = 3.26 in the post-test, thus an average Learning gain of 1.24 (*SD* = 1.15).

Learnings gains are only very marginally higher in the CC condition, yet we conducted a ANOVA to test the significance of this difference. Results, shown in Table 5 below, confirm that this difference is not significant (*F* = 0.28, *p* = .60).

Table 5. Learning Gains by Design Types and Error Orientation using an ANOVA

ANOVA - Learning gains

	Sum of Squares	df	Mean Square	F	p
Design Type	0.419	1	0.419	0.283	0.596
ErrorOrientation	0.303	1	0.303	0.205	0.652
Design Type ✻ ErrorOrientation	0.739	1	0.739	0.499	0.482
Residuals	142.159	96	1.481		

5.2.2 Main effect of Error Orientation

As we can see in Table 3, participants reporting High tolerance for errors showed slightly higher Learning gains (*M* = 1.36, *SD* = 1.43) than those reporting Low tolerance for errors (*M* = 1.23, *SD* = 1.14). As with Design Types, the difference is small, and the ANOVA in Table 5 confirms that Error Orientation does not seem to play a role in Learning Gains (*F* = 0.21, *p* = .65).

5.2.3 Interaction between Design Type and Error Orientation

In Table 4, we can see that Learning gains are higher for participants in the CC condition with a High Tolerance for errors (*M* = 1.55, *SD* = 1.44). Interestingly, participants with High

Tolerance in the CSC condition show the lowest Learning gains (*M* = 1.18, *SD* = 1.47) of all groups. Participants reporting Low Tolerance have very similar results in both CC (*M* = 1.21, *SD* = 1.22) and in CSC (*M* = 1.26, *SD* = 1.07). Yet, we can see in the ANOVA in Table 5 that this interaction has no influence on Learning gains either (*F*= 0.50, *p* = .48).

5.3 Effect on Confidence

5.3.1 Main effect of Design Type

We can see in Table 2 that participants' level of confidence increased in both Design Types, though participants reported slightly higher Confidence levels in the CSC condition (*M* = 1.16, *SD* = 0.72) compared to the CC condition (*M* = 0.94, *SD* = 0.72). We conducted another ANOVA to test the significance of this difference. Results, in Table 6, show us that there is no significant difference between Design Types on Confidence levels (*F* = 0.47, *p* = .49).

Table 6. Confidence Level by Design Types and Error Orientation

ANOVA - Confidence Level

	Sum of Squares	df	Mean Square	F	p
Design Type	0.2461	1	0.2461	0.473	0.493
ErrorOrientation	0.0977	1	0.0977	0.188	0.666
Design Type ✻ ErrorOrientation	0.6061	1	0.6061	1.164	0.283
Residuals	49.9924	96	0.5208		

5.3.2 Main effect of Error Orientation

Participants with High tolerance for errors (*M* = 0.99, *SD* = 0.73) report marginally lower Confidence levels than participants with Low tolerance for errors (*M* = 1.06, *SD* = 0.73), as shown in Table 3. Again, with the ANOVA in Table 6, we observe that there is no significant influence of Error Orientation styles on the Confidence level (*F* = 0.19, *p* = .66).

5.3.3 Interaction between Design Types and Error Orientation

Table 4 reveals that Confidence levels increase in all conditions, the highest being for participants with Low tolerance for errors in CSC condition (*M* = 1.22, *SD* = 0.77), and the lowest being for participant with Low tolerance for errors in CC condition (*M* = 0.91, *SD* = 0.65). People with High tolerance for errors have slightly higher confidence levels in CC (*M* = 1.02, *SD* = 0.96) than in CSC (*M* = 0.96, *SD* = 0.45). Again, the analysis in Table 6 indicates that there is no significant influence from the interaction between Design Types and Error Orientation ($t = 1.16$, $p = .28$).

5.4 Effects on Emotions

We analysed self-reports on emotions to understand what epistemic emotion would emerge during the course, thus also allowing us to measure the success of the induction of confusion in the confusion-based design. As a reminder, we would expect reports of confusion to be higher in the experimental condition (CC) than in the control condition (CSC).

5.4.1 Main effect of Design Type

Results, in Table 2, show similar reports on all 5 emotions throughout Design Types: Participants report high Interest in both CC (*M* = 4.18, *SD* = 0.78) and CSC conditions (*M* = 4.20, *SD* = 0.77). There is also a consistent report of low Frustration and Boredom.

We observe that reports of Confusion are low in both CC and CSC conditions, but slightly higher in the CC condition (*M* = 2.02, *SD* = 0.95) than in the CSC condition (*M* = 1.75, SD = 0.89 in CSC). Similarly, Surprise is higher in CC (*M* = 2.47, *SD* = 0.91) than in the CSC condition (*M* = 2.13, *SD* = 1.06).

Using T-tests, displayed in Table 7, to test the significance of the differences of emotion reports between Design Types, results indicate that the difference for Surprise is significantly higher in the CC ($t = 1.75$, $p < .05$), while Confusion is just below the threshold of significance, thus could be considered as a trend ($t = 1.75$, $p = .08$). However, effect sizes, measured with Cohen's D, are both small for Confusion ($d = .29$) and Surprise ($d = .35$).

Table 7. T-test for Emotions between Design Types

Independent Samples T-Test

		Statistic	df	p		Effect Size
Surprised	Student's t	1.748	98.0	0.042	Cohen's d	0.3495
Confused	Student's t	1.437	98.0	0.077	Cohen's d	0.2874
Bored	Student's t	0.292	98.0	0.385	Cohen's d	0.0584
Frustrated	Student's t	0.845	98.0	0.200	Cohen's d	0.1689
Interested	Student's t	−0.118	98.0	0.547	Cohen's d	−0.0235

Note. H_a $\mu_{CC} > \mu_{CSC}$

5.4.2 Main effect of Error Orientation

Results in Table 3 indicate that like with Design Types, there are only insignificant differences for Emotions, between people with a High tolerance for errors and Low tolerance for errors. Interest is high for both HT (*M* = 4.08, *SD* = 0.73) and LT (*M* = 4.22, *SD* = 0.79), whereas Frustration and Boredom are both low. Neither Surprise, nor Confusion have significant differences, as validated with t-tests.

5.4.3 Interaction between Design Types and Error Orientation

As can be seen in Table 4, there are no significant or relevant differences for our research. Interest is globally similar in all subsets. Regarding Confusion, we see that the difference is in line with the main effects, i.e., mostly on Design Types, but not between Error Orientation. Similarly, Surprise and Frustration are not affected by the interaction, only by the Design Types. Only Boredom seems to be linked to Error Orientation, as it is higher for HT in CC (*M* = 1.71, *SD* = 0.99) and CSC (*M* = 1.59, *SD* = 0.92) than for LT, both in CC (*M* = 1.33, *SD* = 0.70) and in CSC (*M* = 1.31, *SD* = 0.71).

5.5 Correlations between variables

To further our analysis on our data set, we performed an exploratory correlation matrix to detect trends between all dependent variables.

Table 8. Correlation matrix for dependent variables.

Correlation Matrix

		Interested	Confused	Surprised	Frustrated	Bored	Prior interest	Confidence Delta	Learning gains	Error Orientation	Additional Needs
Interested	Spearman's rho	—									
	p-value	—									
Confused	Spearman's rho	−0.039	—								
	p-value	0.699	—								
Surprised	Spearman's rho	0.097	0.675 ***	—							
	p-value	0.335	< .001	—							
Frustrated	Spearman's rho	−0.405 ***	0.575 ***	0.386 ***	—						
	p-value	< .001	< .001	< .001	—						
Bored	Spearman's rho	−0.605 ***	0.432 ***	0.289 **	0.708 ***	—					
	p-value	< .001	< .001	0.004	< .001	—					
Prior interest	Spearman's rho	0.602 ***	−0.122	−0.034	−0.424 ***	−0.539 ***	—				
	p-value	< .001	0.226	0.733	< .001	< .001	—				
Confidence Delta	Spearman's rho	0.347 ***	−0.058	0.126	−0.396 ***	−0.283 **	0.262 **	—			
	p-value	< .001	0.566	0.213	< .001	0.004	0.008	—			
Learning gains	Spearman's rho	0.136	0.172	0.341 ***	0.042	0.022	0.212 *	0.182	—		
	p-value	0.178	0.088	< .001	0.679	0.827	0.034	0.070	—		
Error Orientation	Spearman's rho	−0.305 **	0.145	0.046	0.127	0.267 **	−0.220 *	0.001	0.018	—	
	p-value	0.002	0.151	0.647	0.208	0.007	0.028	0.990	0.856	—	
Additional Needs	Spearman's rho	0.302 **	0.336 ***	0.197 *	0.047	−0.192	0.170	0.089	0.084	0.021	—
	p-value	0.002	< .001	0.049	0.644	0.056	0.090	0.377	0.405	0.838	—

Note. * p < .05, ** p < .01, *** p < .001

Table 8 shows some significant results. In line with major theoretical frameworks, surprise and confusion have a strong positive correlation (r = 0.66, p < .001). Interestingly, confusion also correlates positively with boredom (r = 0.43, p < .001). and frustration (r = 0.58, p < .001).

People who reported a desire for additional resources predominantly reported confusion (r = 0.34, p < .001), as well as interest (r = 0.30, p < .01) and less significantly surprise (r = 0.20, p < .05).

Not surprisingly, reports of high prior interest are significantly correlated with the emotion of Interest (r = 0.60, p < .001) and significantly negatively correlated with Frustration (r = -0.42, p < .001) and Boredom (r = -0.54, p < .001). In the same line, Confidence level is positively correlated with Interest (r = 0.35, p < .001) and negatively with Frustration (r = -0.40, p < .001) and Boredom (r = -0.28, p < .01).

Similarly, Interest correlates negatively with Boredom (r = -0.61, p < .001) and Frustration (r = -0.41, p < .001). Furthermore, reports of high error orientation (i.e., low tolerance for errors) is negatively correlated with Interest (r = -0.31, p < .01) and positively with Boredom (r = 0.27, p < .01).

Regarding our main dependent variables, Learning gains have a strong positive correlation with Surprise (r = 0.34, p < .01) and less significantly with Confusion (r = 0.17, p = .09). Confidence levels on the other hand seem strongly linked to Interest (r = 0.35, p < .001).

6. Discussion

This research aimed to further findings on confusion as a beneficial support for learning by creating instructional design using a confusion-based approach. We tried to identify if confusion-based design could foster learning and in what conditions, specifically by combining it with an Error Orientation style. We also explored if confusion-based instructional designs would influence Confidence levels, as they require deeper processing and would result in spending more time on a topic.

6.1 Hypotheses and results

6.1.1 H1: The feeling of confusion triggers cognitive processes that are beneficial to learning

We tested if confusion would drive the activation of cognitive resources for a deeper learning process. Global results showed that there are no significant differences on Learning gains between Design Types. However, results on self-reports of emotions showed a small, yet significant difference between the two designs. The confusion-based design triggered higher surprise and slightly higher confusion than the classic Direct Instruction Design. This could mean that the confusion-based design was on the right track but missing some elements to really trigger confusion. We postulate that our willingness to create experimental designs as similar as possible to avoid external interferences might have caused a structural problem, that is, our designs are too much alike to be able to produce different kind of emotions. Other suggestions are discussed in the limits.

With our results, we cannot conclude this specific confusion-based instructional design gives stronger learning outcomes than regular instructional design, but it gives us interesting food for thought for other instructional designs, centred around surprise and confusion.

6.1.2 H2: The feeling of confusion fosters learners' confidence

Directly linked to the previous comment, we hypothesised that learners who felt a productive confusion and initiated a deeper processing of information would see a raise in their confidence level. Results show that there are no significant differences between Design Types, nor for Error Orientation. The interaction however seems to hint at the fact that participants with Low tolerance for errors are more likely to report a high confidence level in a classic direct instruction design that in a confusion-based design.

This conclusion is opposed to our hypothesis, yet it seems to shed light on very important nuance: we hypothesized that people would gain confidence through confusion because confusion drives people to invest more in the learning process. What we did not consider before the study though is that this confusion-based learning might educate people to an attitude of critical doubt, that would lead them to know better but also become more cautious in their confidence – akin to the *Valley of Despair* in the Dunning-Kruger Effect, where people lose confidence as they become more competent (Kruger & Dunning, 1999). Again, it is not something that this study aimed at measuring, but that triggers interesting directions for further research in this field.

6.1.3 H3: Confusion-based instructional design is only beneficial if learners have a positive error-orientation style

Results showed that learnings gains were highest for participants with High tolerance for errors while in the confused-based design, and lowest for the same participants in the direct instruction design. This would be in line with the theories we use as frameworks for this research: students with a growth mindset would be driven to a productive confusion, with this specific confusion-based design, and therefore to better learning outcomes, whereas they would not feel the need to process information deeper in a direct instruction class and have smaller learning gains. These differences are not significant though, so we can only highlight these findings a possible further track to follow.

It is worth noting that we tested Confusion-based Instruction Design with Error Orientation Style to identify possible situations in which this design would not be appropriate. We live in a culture that does not promote failure as a means to succeed, as it can be the case in other countries. Therefore, designing with confusion for learners who are afraid of errors and feeling confused could be critically counterproductive. Our results did not show any difference between conditions. However, our sample reported low stress when making errors and a strong perception of errors as a source for learning. This indicates an attitude akin to a growth mindset, not necessarily what would be expected in any population. It would be interesting to test the questionnaire on a bigger sample of students or on other populations to determine the baseline for this test, as we may be wrong in thinking that people are error adverse. Finally, even though the study was fully anonymous, these results may reflect social desirability, as it may feel self-rewarding to report no negative emotions toward errors.

6.1.4 Correlation between emotions and other dependent variables

Despite not seeing significant differences between designs, the analysis of correlations still gives interesting paths to explore. The fact that Surprise is strongly correlated with Confusion is in line with theoretical models, such as D'Mello and Graesser's model (2012). As explained in the next section on limit, our design might not have been complex enough to trigger Confusion, yet incongruent enough to trigger Surprise. People who reported confusion were also more likely to ask for additional resources, thus providing a solid argument that confusion could contribute to trigger deeper thinking and activates cognitive resources.

6.2 Limits

The main limit of this experiment is the low report of confusion in our confusion-based instructional design type (CC). We propose several explanations to better understand and future similar research:

6.2.1 Recruitment

As discussed previously, confusion is a subtle emotion to induce, and it was probably ambitious to expect to successfully induce it in participants who were recruited only for the purpose of this study. One hypothesis is that the length of dialogues in each topic of the course may have been too short to elicit confusion. Studies on confusion have been conducted on longer chapters in courses.

This study was led during an epidemic peak, at the end of winter, before an exam session. Even if participants reported strong motivation for the topic, it is plausible that fatigue and competing factors may have played a role. This leads us to think that these kinds of studies should be conducted on real academic courses or training for which participants have signed-on, thus better controlling the learners' interest and engagement.

6.2.2 Ability to detect and express emotion

Qualitative data collected through informal conversations with the test group as well as a few participants after they had passed the study shed light on an interesting point: some participants did not feel fully equipped to report on their emotions. First, it requires awareness of one's own feelings and emotional states. Several participants expressed similar experiences, in a form similar to: "it was not what I thought / what I have been taught, this surprised me". Using our model, we would have expected them to report confusion and not only surprise. Surprise is the initial state of detecting a discrepancy. It seems that with interest and surprise, people just accepted the new fact presented, as no imbalance was felt to trigger confusion.

Of course, as in any studies using self-reports, social desirability may play a role. In our experiment specifically, it would have been interesting to have additional qualitative data to know if people were culturally more oriented to report on positive than negative emotions, thus not *wanting* to report on confusion rather than not being able to detect its occurrence.

6.2.3 Experience Design

Another assumption is that the context was not relevant enough to elicit confusion. As confusion arises if the appraisal assessment exceeds the threshold for relevance, the participants, benevolent as they may have been, were still not learning the topic based on personal needs or goals. Even though we tried to control this by choosing *First Aid* as a universally subject of interest for most people, competing goals and needs may have been at play on which we had no control. This idea is reinforced by the significant difference on Surprise between the two Design Types. Our design managed to elicit more Surprise than a regular instruction. As Surprise is a direct antecedent of Confusion, participants may have felt Confusion if the course was either more relevant to them, but also simply longer in time or volume (participants read 8 topics and filled in all questionnaires in an average of 18.89 minutes).

6.3 Perspectives

Although this research did not succeed in providing a conclusive framework to design online courses using a confusion-based approach, participants reported high motivation, high interest and demonstrated strong learning gains from the course, which is what instructional designers and teachers seek to achieve. This could mean that the design with dialogues and misconceptions can be an interesting approach in digital courses and that we should explore this approach in further studies.

As confusion is triggered after surprise, it seems that confusion needs more complexity and stronger incongruities to get triggered. Confusion-based design thus seems more adapted to complex learning, for students who have an initial motivation for the topic. It would be interesting to try this type of design with a specific class, for a longer period.

If error orientation styles proved to be a hindrance to regulate confusion, the next steps would be to detect early participants at risk and provide them with a special preparatory course on mindsets and failure as a chance to learn. Generally speaking, providing students and learners with a safe space to explore and test out ideas can also support interest, engagement and curiosity, all of which are positively correlated with learning gains.

Confusion-based instructional design may be an arguable choice of terminology. We used it to refer to impasse-driven theories of learning (VanLehn, 2003), Productive Failure (Kapur, 2016) and specifically to operationalise D'Mello & Graesser's Zone of Optimal Confusion. However, as research on epistemic emotions has been showing, and that we could also detect in this study, learners can be in several emotional states at the same time, and it is possible that what some authors call "productive confusion" (D'Mello & Graesser, 2014) could be called "Flow" for others. The Zone of Optimal Confusion creates a space with enough challenge to be interesting and not boring, and not too difficult to create anxiety. While in this zone, learners feel actively engaged and process information deeper. This is close to the description of flow. Further research to measure the difference between flow and productive confusion could provide interesting insight for instruction design.

To conclude, as technologies in Machine Learning are growing at a rapid pace, these are very interesting times to do research in this field and determine how to detect emotions without getting learners out of their learning and give dynamic, adaptative and individualised support.

7. Acknowledgments

Un merci éternel à Gaëlle pour sa patience quand je vais trop loin et dans tous les sens, sa présence inspirante, ses conseils avisés, et pour nourrir ma curiosité - ou tenter de la canaliser, selon. Je me suis investie dans ce master en grande partie grâce à toi, d'où je tire deux années d'une richesse sans pareil. Merci pour la confiance et l'exemple.

Merci aux Zelda et mon groupe d'écureuils fous pour leur présence joyeuse, les fous rires, les challenges qui poussent à en faire un peu plus mais toujours avec bienveillance, l'envie de partager et la prise de conscience qu'apprendre à plusieurs, oui, ça peut être plus efficace que seule. ☺

Merci à ma Colinette Maud et ma Bibiche Aline pour leur écoute soutenante et leur relecture encourageante, à Julien pour les idées de recadrage dès que je m'égarais trop et à Raúl pour avoir tenté de m'apprendre R en 4 heures.

Merci à tou·te·s mes proches pour leur soutien, et spécialement à mon amoureux pour les mousses aux framboises merveilleuses qui ont permis la finalisation de ce travail dans la béatitude.

8. Bibliography

Arguel, A., Lockyer, L., Kennedy, G., Lodge, J. M., & Pachman, M. (2019). Seeking optimal confusion: a review on epistemic emotion management in interactive digital learning environments. *Interactive Learning Environments*, *27*(2), 200-210.

Bahreini, K., Nadolski, R., & Westera, W. (2016). Towards multimodal emotion recognition in e-learning environments. *Interactive Learning Environments*, *24*(3), 590-605.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: when students" game the system". In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 383-390).

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

Bosch, N., Chen, Y., & D'Mello, S. (2014, June). It's written on your face: detecting affective states from facial expressions while learning computer programming. In *International Conference on Intelligent Tutoring Systems* (pp. 39-44). Springer, Cham.

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.

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. New York: Harper & Row.

D'Mello, S. K. & Graesser, A. C. (2012). Dynamics of affective states during complex learning. Learning and Instruction, 22, 145–157. doi:10.1016/j.learninstruc.2011.10.001

D'Mello, S. K. & Graesser, A. C. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(4), 1-39.

D'Mello, S. K., & Graesser, A. C. (2014a). Feeling, thinking, and computing with affect-aware learning. *The Oxford handbook of affective computing,* 419-434.

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

D'Mello, S. K., Lehman, B., Pekrun, R., & Graesser, A. C. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170. doi:10.1016/j.learninstruc.2012.05.003

de Winter, J. F., & Dodou, D. (2010). Five-point likert items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). *Practical Assessment, Research, and Evaluation*, *15*(1), 11.

Dweck, C. S. (2006). Mindset: The new psychology of success. New York, NY: Random House Incorporated

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological science, 14*(3), 481-496.

Forbes-Riley, K., & Litman, D. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication*, *53*(9-10), 1115-1136.

Gross, J. J. (2008). Emotion regulation. *Handbook of emotions, 3*(3), 497-513.

Huang, X. (2017). Example-based learning: Effects of different types of examples on student performance, cognitive load and self-efficacy in a statistical learning task. *Interactive Learning Environments*, *25*(3), 283-294.

Ismail, M., & Syaiful, L. (2015). Affective assessment in learning using fuzzy logic. In 2015 *IEEE Conference on e-Learning, e-Management and e-Services* (IC3e) (pp. 98-102). IEEE.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289-299.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45-83.

Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In *Proceedings IEEE International Conference on Advanced Learning Technologies* (pp. 43-46). IEEE.

Kruger, J. M., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Lauzier, M. (2011). Impact différentiel des types de modelage et des styles d'orientation des buts sur des indices cognitifs, affectifs et comportementaux liés à l'expérience d'apprentissage.

Lauzier, M., & Mercier, G. (2018). The effect of error orientation, motivation to learn, and social support on training transfer intentions: A moderated mediation model. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration, 35*(3), 419-428.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer publishing company.

Lehman, B., D'Mello, S., & Graesser, A. (2013, July). Who benefits from confusion induction during learning? An individual differences cluster analysis. In *International Conference on Artificial Intelligence in Education* (pp. 51-60). Springer, Berlin, Heidelberg.

Lehman, B., D'Mello, S., Strain, A., Mills, C., Gross, M., Dobbins, A., ... & Graesser, A. (2013). Inducing and tracking confusion with contradictions during complex learning. *International Journal of Artificial Intelligence in Education, 22*(1-2), 85-105.

Linnenbrink-Garcia, L., & Pekrun, R. (2011). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology*, *36*(1), 1-3.

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, *29*(4), 693-715.

Long, Z., Luo, D., Gao, H., & Hu, X. (2019). The Effects of Rejection Sensitivity on Confusion Regulation during Learning in Multiagent Intelligent Tutoring System Environments. In TTW@ AIED (pp. 77-86).

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R. E. (2017). Using multimedia for e-learning. J*ournal of Computer Assisted Learning, 33*(5), 403-423.

Molden, D. C., & Dweck, C. S. (2006). Finding" meaning" in psychology: a lay theories approach to self-regulation, social perception, and social development. *American psychologist, 61*(3), 192.

Muis, K. R., Psaradellis, C., Lajoie, S. P., Di Leo, I., & Chevrier, M. (2015). The role of epistemic emotions in mathematics problem solving. *Contemporary Educational Psychology*, *42*, 172-185.

Muller, D. A. (2008). *Designing effective multimedia for physics education*. Sydney: University of Sydney.

Paas, F. G., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and motor skills*, *79*(1), 419-430.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, *37*(2), 91-105.

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.* (pp. 3-31). American Psychological Association.

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: the Epistemically-Related Emotion Scales. *Cognition and Emotion*, *31*(6), 1268-1276.

Rodrigo, M. M. T., Baker, R. S., Lagud, M. C., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., ... & Viehland, N. J. (2007). Affect and usage choices in simulation problem solving environments. *Frontiers in Artificial Intelligence and Applications*, *158*, 145.

Rodrigo, M., Mercedes, T., & d Baker, R. S. (2011). Comparing the incidence and persistence of learners' affect during interactions with different educational software packages. In *New perspectives on affect and learning technologies* (pp. 183-200). Springer, New York, NY.

Rybowiak, V., Garst, H., Frese, M., & Batinic, B. (1999). Error orientation questionnaire (EOQ): Reliability, validity, and different language equivalence. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *20*(4), 527-547.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion, 23*(7), 1307–1351

Silvia, P.J. (2008). Interest - The curious emotion. *Current directions in psychological science 17*.(1). 57-60

Sinha, T., & Kapur, M. (2021). Robust effects of the efficacy of explicit failure-driven scaffolding in problem-solving prior to instruction: A replication and extension. *Learning and Instruction*, *75*, 101488.

Stein, N. L., & Levine, L. J. (1990). Making sense out of emotion: The representation and use of goal-structured knowledge. *Psychological and biological approaches to emotion*, 45-73.

Sullins, J., Console, K., Denton, R., Henrichson, C., & Barber, S. (2019). Not all confusion is productive: an investigation into confusion induction methods and their impact on learning. *International journal of learning technology*, *14*(4), 288-303.

TIAM-LEE, T. J. Z., & Sumi, K. (2019). Affective Tutoring for Programming Education. In *Proceedings of the Annual Conference of JSAI 33rd (2019)* (pp. 2J1E505-2J1E505). The Japanese Society for Artificial Intelligence.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring?. *Cognition and Instruction*, *21*(3), 209-249.

Vogl, E., Pekrun, R., Murayama, K., Loderer, K., & Schubert, S. (2019). Surprise, curiosity, and confusion promote knowledge exploration: Evidence for robust effects of epistemic emotions. *Frontiers in psychology*, 2474.

Weiner, B. (1972). Attribution theory, achievement motivation, and the educational process. *Review of educational research*, *42*(2), 203-215.

9. Appendix A - Course content

Examples of course content: “Seizure”

	Confusion-based Dialogue (CC)	Direct Instruction Dialogue (CSC)
1		
2		
3		
4		

10. Appendix B - Questions

Examples of questions (same in both pre-test and post-test)

QUESTION “Etouffement” :

“Vous prenez un repas convivial en famille quand, soudain, l’un de vos proches se met à tousser fortement après avoir vraisemblablement avalé une bouchée de travers. Cela ne semble pas passer, il a de la peine à parler et son teint devient rouge. Que faites-vous ?”

- Je l’encourage à tousser et j’observe l’évolution (correct answer)
- Je lui donne des tapes dans le dos en l’encourageant à tousser (distractor close to correct answer)
- Je me positionne derrière lui afin de réaliser la manoeuvre de Heimlich (misconception)
- Je lui donne des tapes dans le dos et me prépare à réaliser la manoeuvre de Heimlich (distractor close to misconception)

QUESTION “Saignement” :

“Vous marchez dans la rue tout en répondant à un texto lorsque vous vous prenez brutalement un lampadaire en plein visage. Votre nez se met à saigner abondamment. Que faites-vous ?”

- Je penche la tête vers l’avant, presse gentiment sur mes narines et observe durant 10 minutes (correct answer)
- Je penche la tête vers l’avant, j’évite de toucher mon nez et observe durant 10 minutes (distractor close to correct answer)
- Je penche la tête vers l’arrière sans toucher à mon nez et j’observe durant 10 minutes (misconception)
- Je penche la tête vers l’arrière, presse gentiment sur mes narines et observe durant 10 minutes (distractor close to misconception)

11. Appendix C - Questionnaires

Error-Orientation Questionnaire

Votre rapport à l'erreur durant l'apprentissage.
Lorsque vous êtes en train d'apprendre, dans quelle mesure les affirmations suivantes s'appliquent à vous :

	Pas du tout	Un peu	Ni l'un, ni l'autre	Beaucoup	Complétement
Je trouve stressant de faire des erreurs	○	○	○	○	○
J'ai souvent peur de faire des erreurs	○	○	○	○	○
Je me sens gêné·e lorsque je fais une erreur	○	○	○	○	○
Si je fais une erreur lorsque j'apprends, je perds mon sang-froid et je me mets en colère	○	○	○	○	○
Lorsque j'apprends, je m'inquiète de faire quelque chose de faux	○	○	○	○	○
Les erreurs m'aident à améliorer mon travail	○	○	○	○	○
Les erreurs me fournissent des informations utiles pour mon apprentissage	○	○	○	○	○
Mes erreurs m'ont aidé à progresser	○	○	○	○	○

Interest for the topic

Dans quelle mesure les énoncés suivants s'appliquent-ils à vous :

	Pas accord	Plutôt pas d'accord	Ni en désaccord, ni d'accord	Plutôt d'accord	Tout à fait d'accord
J'ai aimé en apprendre plus sur les premiers secours	○	○	○	○	○
J'ai trouvé ce cours sur les premiers secours intéressant	○	○	○	○	○
Je trouve que tout ce que m'a appris ce cours sur les premiers secours peut être utile	○	○	○	○	○

Mental Effort

Durant cette formation, vous avez fourni un effort mental...

	Très, très bas							Très, très haut
	1	2	3	4	5	6	7	8	9

Votre réponse