Inside Out: Detecting Learners' Confusion to Improve Interactive Digital Learning Environments

Amaël Arguel [1], Lori Lockyer [1, 2], Ottmar V. Lipp [3], Jason M. Lodge [4], and Gregor Kennedy [4]

[1] Department of Educational Studies, Macquarie University
[2] Faculty of Arts and Social Sciences, University of Technology Sydney
[3] School of Psychology and Speech Pathology, Curtin University
[4] Melbourne Centre for the Study of Higher Education, University of Melbourne

Corresponding Author

Dr Amaël Arguel,
4 First Walk, 931
Department of Educational Studies,
Macquarie University NSW 2109, Australia
amael.arguel@mq.edu.au
Phone: +61 2 9850 8664

Biographies of the Authors

- Dr Amaël Arguel is a psychological scientist specialised in education and multimedia learning. He is a member of the Department of Educational Studies at Macquarie University in Sydney and appointed as a research fellow of the ARC-SRI Science of Learning Research Centre.

- Professor Lori Lockyer researches in the field of educational technology with a focus on learning in school, higher education and professional settings. Lori is a Chief Investigator in the ARC-SRI Science of Learning Research Centre, Honorary Professor at Macquarie University and Dean of the Graduate Research School at the University of Technology Sydney.

- Professor Ottmar V. Lipp is an experimental psychologist whose research is concerned with human learning and the interaction of emotion and cognition. Ottmar uses behavioural and physiological measures in his research. He is a Chief Investigator in the ARC-SRI Science of Learning Research Centre and a Professor in the School of Psychology and Speech Pathology, Curtin University, Perth.

- Dr Jason M. Lodge is a psychological scientist and Senior Lecturer in the Melbourne Centre for the Study of Higher Education. He is also a Senior Research Fellow in the ARC-SRI Science of Learning Research Centre. Jason's research focuses on the application of the learning sciences to higher education and on the ways in which technology is influencing learning.

- Professor Gregor Kennedy undertakes research on the development and use of technology in higher education contexts. He is the Pro Vice-Chancellor of Educational Innovation at the University of Melbourne, is a Professor in the Melbourne Centre for the Study of Higher Education and a Chief Investigator in the ARC-SRI Science of Learning Research Centre.

Abstract

Confusion is an emotion that is likely to occur while learning complex information. This emotion can be beneficial to learners in that it can foster engagement, leading to deeper understanding. However, if learners fail to resolve confusion, its effect can be detrimental to learning. Such detrimental learning experiences are particularly concerning within digital learning environments, where a teacher is not physically present to monitor learner engagement and adapt the learning experience accordingly. However, with better information about a learner's emotion and behaviour, it is possible to improve the design of interactive digital learning environments (IDLE) in promoting productive confusion but also in preventing overwhelming confusion. This article reviews different methodological approaches for detecting confusion, such as self-report, behavioural and physiological measures, and discusses their implications within the theoretical framework of a zone of optimal confusion. The specificities of several methodologies and their potential application in IDLEs are discussed.

Keywords: confusion, interactive learning environments, interface design, learning, emotion

Introduction

Interactive digital learning environments (IDLE) are now ubiquitous in formal and lifelong learning contexts. These environments often provide access to a vast amount of comprehensive information, and a multitude of learning tasks that may be more or less structured. Moreover, depending on their implementation, they often assume learners will work alone with limited or sporadic access to a teacher or facilitator. The complexity of learning activities in these environments may produce an emotional response in learners that may support or hinder their learning. For example, unexpected feedback can challenge students and interrupt the flow of a learning sequence. In the classroom teachers are most of the time able to assess the nature of the emotional states their students experience and the intensity of these states. Teachers can then react accordingly, for example by slowing down the pace of activities if students seem perplexed, or inversely by offering a challenge to solve if students seem disengaged from the learning task. In IDLEs, the emotional states of students are not easily monitored yet, despite a call for systems to be adaptive to emotional responses.

The emerging field of *affective computing* focuses specifically on the influence of emotions in human-computer interactions, including learning interactions. This field considers ways of detecting users' emotions and also the simulation of affective responses by computer systems (see Calvo, D'Mello, Gratch, & Kappas, 2014). In digital contexts, some basic emotions such as anger, disgust, happiness, sadness, and less obviously, surprise (Ekman, 1992) do not seem to be as relevant as is a set of more complex, academic emotions such as confusion, boredom, frustration, and flow (Pekrun, Goetz, Titz, & Perry, 2002; Pekrun & Stephens, 2012). This article focuses on one of these more complex emotions in IDLEs: confusion.

Confusion is an emotion that is likely to occur while acquiring complex knowledge. This emotion can be beneficial to learners in that it assists in fostering engagement and can support the development of a deeper understanding. However, if learners fail to resolve confusion after a period of time, its effect can be detrimental to learning (Rodrigo et al., 2009). In these instances, learners can experience negative feelings, such as frustration and then boredom, before giving up. Such detrimental learning experiences are particularly concerning within IDLEs if a teacher or the learning environment itself is not able to identify or monitor learner engagement and adapt the learning experience accordingly.

With better information about learners' emotions, it may be possible to design vthat effectively engage learners and/or support them with feedback or self-regulation strategies. However, individual differences among learners mean there are no universal rules about when and how to intervene on the basis of emotional responses to the learning experience. Hence, real-time and individual detection of confusion during learning could provide crucial information and improve the quality of adapting and scaffolding learning pathways.

In this article, we firstly present a conceptual framework of confusion in IDLEs. We then review relevant methodological approaches for detecting confusion and provide insights for real-world applications of confusion detection.

Understanding Confusion in Learning

Among academic emotions, confusion is particularly interesting because of its complex nature. Confusion is an emotion that is thought to occur spontaneously during complex learning tasks (D'Mello, Lehman, Pekrun, & Graesser, 2014; Lipson, 1992). A remarkable characteristic of confusion is that its impact on learning outcomes can be either beneficial or detrimental, depending of how confusion is handled by learners. It is hence important for educators to understand its process, maintenance and resolution. Because of its prevalence in learning situations, the present article aims to address the topic of detecting learners' confusion with a particular focus on IDLEs.

A Conceptual Framework of Confusion in Interactive Digital Learning Environments

In learning situations, a learner can experience confusion as an affective response to the cognitive processing of information, and as such has been considered an *epistemic emotion* (Pekrun, 2006). This concept of confusion is used in this article because it reflects situations in which a learner may respond to new information which might be inconsistent with their existing knowledge structures (D'Mello et al., 2014). This experience is particularly relevant to IDLEs in which learners may have access to diverse information.

When attempting to learn new material, learners may detect some inconsistencies within the information presented, or between the information presented and their own prior knowledge. These inconsistencies lead to a *cognitive disequilibrium* or *impasse*, which can also be brought about by an unexpected novel issue. The cognitive disequilibrium can for example be produced with what has been referred to as a 'breakdown scenario' that describes when the behaviour of a system (e.g., a key that does not turn in a lock anymore) is abnormal or does not go as expected (D'Mello & Graesser, 2012; Pekrun & Stephens, 2012). In this case, the cognitive disequilibrium can be caused by the perception of an unexpected response from a system (i.e., the system ceases to function properly or breaks down), which would be inconsistent with the prior understanding that learners have about it.

Unexpected feedback to a learner's response within a task can also be a source of cognitive disequilibrium. For example, a surprising outcome of an action (e.g., being told you have the 'wrong answer' as feedback in a quiz) or a disagreement with an intelligent tutoring agent (D'Mello & Graesser, 2006) can lead to a similar experience of disequilibrium. All these types of impasses are likely to provoke confusion, which is interpreted as the affective signature of a cognitive disequilibrium (Craig, Graesser, Sullins, & Gholson, 2004a; Lehman, D'Mello, & Graesser, 2012; Sullins & Graesser, 2014). The results from several studies suggest that achievement of understanding of complex learning sequences is typically linked with impasses and consequently with confusion (D'Mello & Graesser, 2012; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). Confusion can be regarded as a normal stage of the learning process and to understand associated benefits and risks is crucial for educators when designing effective IDLEs.

The Benefits and the Risks Associated With Confusion

Confusion can potentially be beneficial or detrimental in the learning process. This notion is supported by research focused on the dynamics of emotions within IDLEs, particularly intelligent tutoring systems. In a series of studies, researchers have asked participants to report the emotion that described best their current feeling at regular intervals (D'Mello & Graesser, 2012; D'Mello et al., 2014). Results from these studies have shown that: (1) emotions evoked during learning tasks are transient and (2) that the transitions between emotions follow typical patterns. The resulting model based on these observed patterns, called the *model of affect dynamics*, describes articulations between four emotional states:

Flow/Engagement ⇄ *Confusion* ⇄ *Frustration* ⇄ *Boredom*

This model highlights that significant transitions from one emotion to another are only observed between immediate neighbours. For example, the immediate transition from flow to frustration, or from confusion to boredom, is observed only infrequently.

Progress in complex learning environments seems to be associated with experiencing confusion at some stage during the learning task. Indeed, because of the tight links between cognitive processes and confusion, the latter is likely to be present during any challenging learning sequence (Lehman et al., 2012). This can be explained by the physiological implications of disequilibrium for the learner. Mandler's *discrepancy theory* (Mandler, 1984) likens cognitive disequilibrium to an interruption that breaks the continuous flow of acquisition of information in a learning sequence. An interruption can potentially produce a change of mood and physiological responses of learners such as heart rate and electrodermal activity (Macdowell & Mandler, 1989). These physiological changes are the indicators of a general increase of emotional arousal, which can be beneficial in promoting a greater cognitive engagement in the learning task. In such situations confusion is considered constructive because it leads to positive learning outcomes. That is learners are able to *resolve* confusion by engaging deeper cognitive activities (D'Mello & Graesser, 2012). Empirical evidence consistent with a positive effect of confusion on learning performances has been collected in several studies (D'Mello & Graesser, 2014; D'Mello et al., 2014; Graesser & D'Mello, 2012; Lehman et al., 2012; Lehman, D'Mello, & Graesser, 2013). Therefore, experiencing confusion during learning is not necessarily unwelcome, and pedagogical interventions designed to induce confusion can sometimes be an effective way to facilitate learning.

However, confusion can also be associated with negative learning experiences. According to the model of affect dynamics, confusion can lead to a state of frustration that eventually can lead to boredom (D'Mello & Graesser, 2012). When learners fail to resolve their confusion in IDLEs, where they maybe have less scaffolding or support, it can be detrimental, leading to frustration or boredom. This scenario is likely to generate a negative experience for learners and might contribute to them giving up on the learning session altogether (Baker, D'Mello, Rodrigo, & Graesser, 2010; D'Mello et al., 2014). In this case, confusion can be retrospectively labelled as 'non-constructive', or 'non-productive', because of its detrimental effect on learning outcomes (Arguel & Lane, 2015). Hence, it may be reasonable to consider that detecting different types of confusion – constructive and non-constructive – would not be informative since the confusion experienced during learning is the same regardless of its outcome from the learner's point of view.

The effect – positive or negative – produced by confusion depends on whether a learner is able to resolve the confusion by engaging in in-depth processing of the source of confusion: a cognitive disequilibrium (Lehman & Graesser, 2015). The resolution of the cognitive disequilibrium is crucial: it leads the learner from confusion to an affective state of engagement/flow, whereas a failure to resolve might lead to frustration (D'Mello & Graesser, 2012). Because it seems impossible to distinguish constructive confusion from non-constructive confusion when confusion occurs, there is a need for mechanism or method to accurately detect confusion and its characteristics such as duration or intensity, which could allow the prediction of future positive or negative outcomes.

The Zone of Optimal Confusion

As described above, there are good reasons to think that confusion during learning can be beneficial when it produces a response by learners that engages them in deeper information processing. But it can also be detrimental when the resolution of the impasse is not successful. This area between a minimum level and a maximum level of confusion can be conceptualised as a *zone of optimal confusion* (D'Mello et al., 2014; Graesser, 2011). The zone of optimal confusion is related to individual characteristics of learners, such as their prior knowledge regarding the topic of the learning (Lodge & Kennedy, 2015). According to this conception, when learners detect an impasse during a learning sequence, they can move from a state of *cognitive equilibrium* to a state of *cognitive disequilibrium*. As a result they can begin to lose understanding of the learning material and start to experience confusion. The boundary between these two states can be modelled with the threshold T_a. From there, the learner can either resolve the impasse and return to an engaged learning, or they can stay stuck if they fail to revolve the impasse. From here, learners would move beyond a second threshold (T_b) and progress from confusion to frustration (see Figure 1).

– Insert Figure 1 here –

IDLEs can be designed to support learners to positively navigate the zone of optimal confusion through the implementation of features that address two types of interventions (Arguel & Lane, 2015). Firstly, inducing confusion in order to cross the threshold T_a can be a useful technique to engage learners in a difficult learning task. This intervention can be achieved, for example, by highlighting some contradictions in the learning material, or by causing a cognitive conflict between new pieces of information and learner’s naïve conception (Limón, 2001). This is expected to induce a temporary confusion, which once resolved, should foster engagement in the learning activity. However, if learners fail to resolve their confusion, they are likely to drift towards frustration, passing the threshold T_b. To prevent this unwanted event, the learning environment can offer a second type of intervention, designed to manage the level of confusion below T_b. For example, some guidance can be provided in the form of adaptive feedback messages, or by the induction of self-regulation strategies (Butler & Winne, 1995; Narciss, 2004). The main problem with interventions aiming to control trajectories of learners within the zone of optimal confusion is that the locations of the thresholds can be highly variable from one individual to another. This means that it is important to explore techniques that can individually detect the confusion of learners.

Inter-Subject Variability and the Necessity of Detecting Confusion

Around the zone of optimal confusion, the thresholds T_a and T_b are not fixed and their positions depend of several individual factors. For example, factors such as age, motivation, personality, confidence, level of prior knowledge, and learning approach are likely to have a moderating effect on the impact of confusion over learning outcomes (Lehman et al., 2013; Sullins & Graesser, 2014). Consequently, it does not seem reasonable to design learning situations in which the boundaries of confusion would be the same for all participants.

Interventions aiming to induce or to manage the level of confusion, as described above, require accurate information about the learner's real-time experience of confusion. In the face-to-face classroom settings, dealing with confusion is less problematic because teachers can observe physical and verbal cues that allow them to detect when students are experiencing confusion (Goleman, 1995). Hence, teachers may adjust learning activities to accommodate students' different levels of confusion. In IDLEs, the individual learners' level of confusion, and moreover, differences between learners' confusion are more difficult to assess and thus may be overlooked. Consequently, the learning activity may be less adaptive than in a face-to-face learning experience facilitated by a teacher and it could be more difficult to avoid the negative effects of confusion. Implementing systems and strategies in IDLEs, which could detect and control confusion could reduce the risks of frustration which is likely to happen when confusion is left unresolved (see Calvo et al., 2014). For this reason, numerous studies have attempted to develop methods to detect confusion in learning environments.

The Detection of Confusion

Self-Report

If confusion is an emotion, it is possible that those who experience it would be able to report it. Hence, the simplest method of measurement is to ask participants to report their level of confusion during or after learning tasks. In some studies, confusion scores were reported from a binary choice (i.e., 0 = not confused, 1 = confused) at regular intervals during a learning session (Lehman et al., 2012). In others studies, a finer measurement was obtained using Likert-type scales, on which participants rated their level of experienced confusion, for example from 1 to 6 (D'Mello & Graesser, 2014), or from 0 to 10 (Lehman et al., 2012). Another possible method consists of asking participants to choose from a list of emotions the one they felt the most appropriate to describe their emotional experience at specific points of the learning session (Baker et al., 2010; D'Mello & Graesser, 2014). Some experimenters have also asked participants to orally communicate their feelings during the interaction with a learning environment, in an *emote-aloud* protocol (D'Mello & Graesser, 2006; Sullins & Graesser, 2014). This was done synchronously at regular intervals (Baker et al., 2010), or retrospectively using video recordings of participants' faces to cue recall and to rate their emotions after a learning task (D'Mello & Graesser, 2014). But despite its apparent ease-of-use, self-reporting of emotional states can be a problematic data collection method.

Self-reporting of emotions is understandably sub-optimal when the execution of this additional task interferes with the primary learning task and affects learning performance. Another issue with self-reporting is the lack of sensitivity, possibly due to social and cognitive biases such as honesty and willingness to report confusion. Moreover, it has been shown that some emotional intelligence is required from participants to be able to correctly label their emotions (V. Allen, MacCann, & Matthews, 2014; D'Mello & Graesser, 2014; Goleman, 1995). Furthermore, it seems that some elements of emotional behaviour are difficult for learners to consciously process (Calvo & D'Mello, 2010). For all these reasons, attempts to include different ways of measuring confusion have been employed in many studies. For example, patterns of observed behaviours from learners or physiological responses to confusion have been studied in an attempt to obtain objective and reliable indicators.

Behavioural Responses

The effort made by learners while resolving a cognitive disequilibrium can sometime be visible from specific facial expressions and more generally from some changes observable in behaviour. When engaged in an interactive learning task such effort may be observed from changes in postures, conversational cues, on-screen visual exploration, and from the interactions of learners with the interface. These possible indicators of confusion are addressed in the following sections.

Facial expressions.

Facial expressions represent an obvious way to detect emotions from others. Early research considered the expression of emotions from an evolutionary point of view, defining their origins from pragmatic responses to stimuli, which would be then associated with a broader range of situations (Darwin, Ekman, & Prodger, 1998). For example, teachers are able to interpret facial expressions in order to detect when their students are confused (Apps, Lesage, & Ramnani, 2015; Lipson, 1992). Facial expressions are hence natural candidates to serve as indicators to determine emotions from other persons.

Some studies have attempted to objectively identify the facial expression of confusion as an alternative to self-report approaches. The facial action coding system (FACS) is a tool designed to assist the detection of emotions from participants' faces, using an observation grid to break down expressions into several *action units* (Cohn, Ambadar, & Ekman, 2007; Craig, D'Mello, Witherspoon, & Graesser, 2008). Although the FACS was initially created to detect only basic emotions (i.e., happiness, sadness, surprise, disgust, anger, and fear), it seems possible to extend its application to educational settings and to detect confusion as an academic emotion (Sullins & Graesser, 2014). This would however require determining which patterns of action units are involved in facial expressions specifically related to confusion.

In IDLEs, the implementation of systems able to automatically detect confusion from video capture of facial expressions could provide a helpful indication of how a learner understands a particular task (Shan & Braspenning, 2010). Promising results have already been observed with such systems. For example the *Computer Expression Recognition Toolbox* (CERT) is a frame-by-frame tracker of facial expressions based on specific facial cues such as eyebrows, eyelids, and the mouth of learners, captured by a standard webcam (Grafsgaard, Wiggins, Boyer, Wiebe, & Lester, 2013; Littlewort et al., 2011). This system was developed and validated by human judges using FACS to determine the activation of specific facial action units for the detection of particular emotions. For confusion, the action unit 4 (AU4, 'Brow Lowerer') was empirically identified to be the best indicator (Craig, Graesser, Sullins,

& Gholson, 2004b; Grafsgaard, Boyer, & Lester, 2011). Recently, an automatic system for tracking confusion has been tested to detect potential risks taken by elderly users when misunderstanding medical information from instructional videos (Postma-Nilsenová, Postma, & Tates, 2015). The results showed that the automatic detection software, which was based on CERT, was more precise than human observers, in both accuracy and sensitivity, demonstrating the promise of this type of technology for the future.

Facial electromyography.

Facial electromyography (EMG) is another technique that has been used to objectively identify facial expressions of confusion and goes beyond the visual observation of the face. An electromyogram measures the electric activity of contracting muscles with electrodes placed on the surface of the skin. Since facial expressions are produced by the activation of groups of muscles, specific to each emotion, using a facial EMG is assumed to reflect expressions, even when they are barely visible on the learner's face (Bradley & Lang, 2000; Dimberg, 1990; Levenson, Ekman, & Friesen, 1990). The detection of confusion is typically performed by measuring some activation patterns of the muscles surrounding eyes and mouth of participants: the *corrugator supercilii*, the *zygomaticus major*, and the *depressor anguli zoris* muscles (Durso, Geldbach, & Corballis, 2012; Rozin & Cohen, 2003; Sato, Fujimura, & Suzuki, 2008). Encouraging early results from the use of facial EMG have been already reported.

Measurements from facial EMG have been used for detecting confusion in the field of aviation psychology (Durso et al., 2012). Confusion was equated with the *loss of situation awareness* as experienced by aircraft pilots (Durso & Gronlund, 1999). The study, carried out in a flight simulator, produced interesting findings: data from facial EMG allowed detection of confusion even when changes in facial expressions were not visible. Facial EMG may represent a promising technique in terms of sensitivity for the measurement of confusion (Huang, Chen, & Chung, 2004) or the valence (i.e., positive vs. negative) of a broader range of emotions (Larsen, Norris, & Cacioppo, 2003; Mandryk & Atkins, 2007).

The use of facial EMG to detect confusion is mildly intrusive because learners must wear several surface electrodes fixed on the face. This requires a considerable equipment setup, and the technique produces large sets of data that make analyses difficult to perform (Healey, 2014). Hence, facial EMG is currently only possible in lab-based research rather than in naturalistic studies or real world learning situations. However, potential findings resulting from lab-based work can be valuable for the validation of visual detection of expressions and for the development of predictive models of confusion during learning.

Postures and conversational cues.

Beyond facial expressions, it is plausible that confusion may be expressed in other ways. Some evidence suggests that the observation of the learners' body language may provide indications about their affective emotional arousal (D'Mello & Graesser, 2007). Moreover, because body motions are relatively unintentional compared with facial expressions, identifying emotions based on body cues could consequently be less biased by social editing (Calvo & D'Mello, 2010). This offers the advantage of increasing the validity in this type of indicator of emotions. This approach to detecting specific emotions, such as confusion, has been used in studies employing observation grids and trained judges. For example, judges were able to code manifestations of confusion during a learning activity by reporting participants' behaviours such as scratching their head and changes of upper body and head position, in addition to some non-linguistic vocal expressions of confusion (e.g., "Huh?") (Baker et al., 2010). However, this technique requires real-time observers and is hence difficult to implement in digital learning environments.

In order to be deployed in IDLEs, independent systems for the detection of emotions from body postures or verbal cues need to be developed. Some research has attempted to quantify some behaviours in order to automatically measure confusion without any coding from observers (Kapoor, Mota, & Picard, 2001). For example, in a learning task on a computer, a video signal from a web camera was used to measure head poses and hand movements, in order to detect emotions such as frustration, engagement, distraction, and boredom (Asteriadis, Tzouveli, Karpouzis, & Kollias, 2009). In another study, a wireless sensor tracking overall body movements was used to distinguish several emotional states including confusion and boredom, in order to provide automatic feedback to students in an IDLE (Caballe et al., 2014). Another possible methodology for measuring body postures is to use pressure sensor sheets placed on the seat and the backrest of the chair of participants (Tan, Lu, & Pentland, 1997). This system is able to detect every change of body posture that can be used as an emotion indicator. For example, when learners move their body forward to get closer to the computer screen they may be confused, or inversely, when they lean on the backrest, that can possibly reflect a disengagement from the learning task. Data from a similar chair sensitive to body postures has been also used to detect negative and strong emotions like confusion (D'Mello & Graesser, 2012). However, in D'Mello and Graesser's (2012) study, the sensitivity of the measuring instrument was significantly improved with the inclusion of cues extracted from participants' dialogue.

Conversational cues.

Another way to detect "natural" expressions of confusion may be found in dialogue that learners can have, for example, with an intelligent tutoring agent. The agent can be a virtual representation of a character who guides students in a learning sequence using natural language (Soliman & Guetl, 2010). Indeed, when natural language is used to interact with the agent, some indicators of confusion can possibly be extracted from conversational cues. Of course, this kind of indicator can only be collected from learning situations that involve a conversation between learners and the system. However, as the ability of intelligent tutoring systems to interact with learners using natural language is improving constantly, using indicators from speech of learners could be a viable solution to detect confusion in IDLEs (D'Mello, Craig, Witherspoon, Mcdaniel, & Graesser, 2008). Despite the promise of natural language processing in this context, this method also presents the obvious limitation of being implementable only in learning tasks that include verbal interactions, which obviously reduces the range of applications.

Visual exploration.

Another opportunity to detect learners' emotions is in the examination of their visual exploration through eye tracking techniques. Eye tracking is a method that dynamically measures the location of gaze on a scene or a computer screen by using, in most cases, the reflection of an infrared source of light on the cornea. The eye tracker identifies the locations and durations of visual fixations on the screen, which are thought to reflect the allocation of attention. This methodology is generally useful to study the strategies of visual exploration that participants employ during learning. Some researchers have also hinted that eye tracking could be helpful to assess the level of confusion experienced by learners. Data collected from eye tracking cannot provide, by themselves, direct measures of confusion, but might nevertheless provide cues to infer it. This can be done because confusion is the emotional expression of a cognitive disequilibrium, and cognitive disequilibrium during learning can be reflected in changes in the visual exploration of the presented information (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005). In certain learning situations, such as the solution of visual problems, the visual exploration strategies of students captured by eye tracking can also provide some operational cues for an early detection of confusion (Pachman, Arguel, Lockyer, Kennedy, & Lodge, in press). Similarly, gaze directions can be used as indicators of frustration and engagement with the learning task (Asteriadis et al., 2009). Others researchers have pointed out a possible link between mental workload and eye movements, which would tend to become more restricted when the workload is high (May, Kennedy, Williams, Dunlap, & Brannan, 1990). Based on a similar hypothesis, a study has explored the association between eye movements patterns (i.e., fixations duration, total number of fixations) and subjective measures of confusion while performing task on a device emulated by a computer (DeLucia, Preddy, Derby, Tharanathan, & Putrevu, 2014). The results revealed some positive correlations between these variables, highlighting the potential of eye tracking to reflect levels of confusion while using a device.

Eye tracking seems to be a useful methodology to collect information on focus of attention, motivation, and emotional status of learners. Real-time eye tracking could be used as a predictor of the confusion called retrospectively non-constructive, contributing to timely interventions in IDLEs (Hua Wang, Chignell, & Ishizuka, 2006). However, in real-world IDLEs, one of the current challenges for implementing eye tracking-based interventions is the collection of usable data from remote students, due to the low quality of inexpensive solutions and the high cost of good quality equipment. For this reason, in real-world applications, it would be more practical to avoid relying on techniques that involve additional equipment besides the computers that students use.

Learner-computer interaction analysis.

Detecting confusion, and more generally emotions, from learners' behaviours may also be achieved from data automatically collected by the IDLE. An emerging field of research called *learning analytics* deals with the collection and analysis of data that learners produce when engaging with a digital learning platform (Siemens, 2013). In its simplest form, learning analytics can focus on the on-screen behaviours of learners, for example mouse pointer movements, clicks and scrolls, the amount of time spent on webpages, opening of hyperlinks, etc. More elaborated indicators can also be collected from specific task completion such as the responses given in assessment questionnaires, social interactions in discussion forums, production of annotations, and the like. These interactions may allow the quantification of behaviours that participants manifest while learning with a computer, and may be potentially used to derive distinct behavioural patterns that align with and differentiate emotional states of participants engaged with a computerised learning environment. For example, in a study conducted with school students, researchers triangulated log files generated from a web-based system with the coding data of behavioural observations made at the same time (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013). The log files were reporting actions from learners such as requesting a hint, giving a correct or wrong answer, taking a pause after an answer, spending time with help, etc. The data collected from the observation of students' behaviours allowed the training of classifiers based on machine learning algorithms and afforded the automatic classification of learner's emotions (i.e., boredom, frustration, engagement, confusion, off-task, and gaming). This approach offers the advantage that the collection of data from students' interaction behaviours during learning is possible within most IDLEs. Learner-computer interaction data can be either analysed after learning, for research purposes, or in real-time during the learning session. In the latter case, predefined algorithms may trigger actions to manage learners' confusion accordingly.

The principle of learner-computer interaction analysis is to collect cues from behaviours while the learner engages with an interface that offers possibilities of interaction. These cues, such as recognisable patterns of actions, can be empirically linked to cognitive states and emotions such as confusion. However, there is another possibility for detecting confusion: rather than tracking subtle manifestation of confusion from the learning activity itself, learning interfaces can be designed to allow learners to self-report their emotions during an activity. For example, *Emotcontrol* is a tool developed to encourage learners to provide affective feedback in IDLEs (Feidakis et al., 2014; Feidakis, Daradoumis, Caballé, & Conesa, 2013). Emotcontrol consists of an on-screen window displaying a clickable wheel of emotions, associated with colours, and text fields in which students can indicate their feelings. Although this method is intrinsically related to the techniques of self-report previously mentioned, the fact of being embedded in the IDLE allows additional features. For example, the data are exploited in real-time to trigger interventions consisting of the scaffolding from an affective virtual agent. Another functionality of the system is to produce visualisations of data in the form of individual emotional cartography that can be used by instructors and/or peer students. Simpler solutions for collecting emotion awareness data are also possible, offering the advantage of being less disruptive to learning tasks. For example, the provision of an "I am confused" button in a interface gives accurate indications about when learners get into trouble with the learning material (Conati, Hoque, Toker, & Steichen, 2013). Preliminary results of the latter study showed that participants were interested and can use the confusion button efficiently during a task that involves the use of an interactive visualisation supportive tool for decision-making.

The techniques presented so far are based on the observation and measurement of learners' behaviours. Because the subject is the detection of emotions, such as confusion during learning, it is conceivable that learning activities elicit changes in students' emotional arousal and that these changes could be captured by monitoring particular physiological responses.

Physiological Responses

Beyond regarding the behavioural expressions of emotions, another possible approach would be to consider their effects on peripheral physiology. Some emotions including confusion are linked to an increase of general arousal that can be observed from changes in physiological responses (Healey, 2014). In this section, the most commonly used physiological measurements for detecting emotions are presented.

Electrodermal activity.

The electrodermal activity (EDA), also called galvanic skin response (GSR) or skin conductance, is the measurement of a change of the electric conductivity of the skin. Electrodes are placed generally on the palm of the hand or the tips of fingers, and variations of EDA are recorded during the completion of cognitive tasks. These variations of EDA are the reflection of a change of the activity of eccrine sweat glands and there is a consistent body of evidence that sweat secretion is correlated with emotional arousal (Fowles, 1980; van Dooren & Janssen, 2012). EDA has been shown to have a relatively good sensitivity in detecting strong emotions (i.e., high arousal) but also a low ability to discriminate their valence, that is, between positive and negative emotions (Bradley & Lang, 2000). However, the development of a protocol targeting specific learning experiences can allow the linkage of an EDA signal to identify emotions resulting from cognitive processes. For example, Pecchinenda and Smith (1996) manipulated the difficulty of a problem-solving task consisting of the completion of anagrams and found a relationship between EDA and participants' level of engagement in the task. A drop in skin conductance was observed when learners were confronted with extremely difficult problems, which led them to abandon any hope of success and consequently disengage from the task. Even if not discussed in these terms in the study, this situation can be interpreted as an instance of unresolved confusion producing an exit point from the zone of optimal confusion, which is required for the successful solving of the problem.

EDA represents a promising technique to detect emotions during learning, such as engagement, boredom, or confusion (Shen, Wang, & Shen, 2009). Moreover, the recent development of wireless and wearable devices able to detect EDA could also be a facilitating factor for integrating this type of measurement in protocols designed to detect confusion.

Heart rate and heart rate variability.

Similar to EDA, heart rate and heart rate variability (HRV) can be used as physiological indicators of changes of emotional arousal and/or level of workload during learning or task achievement (Aasman, Mulder, & Mulder, 1987; Paas, Van Merriënboer, & Adam, 1994; Tattersall & Hockey, 1995). Heart rate reflects simple changes in cardiac chronotropy, increases and decreases in the number of heart beat per minute, whereas HRV captures the extent of variation within the signal, usually over a period of several minutes. When assessed in the frequency domain HRV can provide estimates of parasympathetic and sympathetic drive to the heart (for a review see J. J. Allen, Chambers, & Towers, 2007) Because it is likely that learners' confusion triggers physiological changes, confusion should produce some visible variations in both EDA and cardiac signals. Moreover, EDA and HR/HRV can be considered as objective ways of measuring confusion because they are the result of mostly unconscious and uncontrolled reactions of the autonomic nervous system (Bradley & Lang, 2000). However, the small effect size produced, as well as the difficulty of collecting a good signal in participants who work on a keyboard while interacting with a computer, is likely to lower the sensitivity and the accuracy of these physiological measurements (Paas et al., 1994). Nevertheless, recording HR/HRV is relatively easy when compared with other physiological indicators and the collected signal is quite robust to noise (Healey, 2014). Consequently, HRV could be a convenient measurement in laboratory-setting experiments, but other factors such as age, posture, level of fitness, and circadian cycle can modulate HRV. In addition, another limitation would be the relatively long recording intervals required for reliable estimates of HRV, which can limit its utility as an indicator of confusion.

Brain imaging.

Neuroimaging methods offer the possibility of detecting emotions from the observation of the working brain. Despite the cost and the complexity of experimental protocols involving brain imaging, the collected signals can be valuable in mapping the confusion events that occur during learning (Calvo & D'Mello, 2010). For example, functional magnetic resonance imaging (fMRI), a method that tracks changes in brain blood flow, has been used to observe changes in the neural activity of the posterior medial frontal cortex when learners were confronted with unexpected feedback, which was likely to cause confusion (Hester, Barre, Murphy, Silk, & Mattingley, 2008). Another non-invasive way to monitor brain activity is the Electroencephalogram (EEG). This technique involves measuring electrical activity of the brain with a set of surface electrodes attached to the scalp of learners. EEG equipment is easier to operate, considerably cheaper, and offers better mobility than do fMRI scanners and past research suggested that EEG could be used to detect epistemic emotions such as frustration (Marosi et al., 2002). Moreover, a study on learning from instructional videos in a Massive Open Online Course (MOOC) showed that using EEG to detect students' confusion was as efficient as trained human observers who rated confusion by monitoring the body language of students (Haohan Wang et al., 2013). In this study 20 online education videos assumed to be either confusing or not confusing (50:50 distribution ratio) were presented to university students. An EEG signal was measured with a single-channel system and electrodes placed over the frontal lobe of participants. Despite the positive, though weak, results of EEG to detect confusion, the authors also highlighted some limitations of using this measure in real-world setting. In particular, they mentioned the cost, the problems with confounding effects of other processes such as mental effort, and privacy concerns from participants who may be reluctant to share their brain activity data.

Although brain imaging is a promising methodology for research on the neuroscience of emotions, this approach has yielded to date only limited success in discriminating emotions (Kemp, Krygier, & Harmon-Jones, 2014). Even if sometimes of use in laboratory-based studies, the difficulty of implementing brain imaging in real-world learning situations would probably hinder the application of this technology for the detection of learners' confusion. Moreover, inferences from findings of brain imaging studies would probably be difficult to apply to real-world learning sessions because of the dissimilarity between the two situations.

Pupillometry.

Many eye tracker systems capable of measuring pupil dilation with reasonable accuracy are now available, allowing the integration of this measure into experimental protocols. The change of the size of the pupil diameter has been studied in numerous domains and has been assumed to reflect some emotional states like, for example, the stress of guilty participants in deception detection studies (Elkins, Zafeiriou, Pantic, & Burgoon, 2015; Lubow & Fein, 1996). There is also some evidence that pupil dilatation could be an indicator of the level of cognitive load during learning; a larger pupil size reflecting a higher workload (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). In the domain of emotion detection, it may be that an increase in pupil diameter would be a valid indicator of emotional arousal. In a study measuring the pupil responses of participants to presentations of emotional pictures, an increase of pupil diameter when viewing emotional pictures (pleasant or unpleasant) has been observed relative to neutral pictures (Bradley, Miccoli, Escrig, & Lang, 2008). In another study, changes of pupil dilation during a learning task have been be used as a physiological indicator of surprise (Preuschoff, ’t Hart, & Einhäuser, 2011). Similar findings have been reported for the detection of confusion based on pupil size changes (Umemuro & Yamashita, 2003). Application of this measurement technique is promising, in particular because the progress of eye trackers available in the market might offer a wider range of devices capable of measuring pupil dilation in the future.

Discussion and Conclusion

Confusion is an emotion that often occurs during the learning of complex material and can lead to negative effects when not resolved within a reasonable period of time. Therefore, timely interventions that aim to help learners resolve their confusion can help support quality learning. When a state of confusion is detected, interventions may consist of providing students with, for example, adaptive feedback or specific guidance. In the classroom, face-to-face interactions between students and teachers can allow teachers to recognise confusion and to modify and adapt their teaching approaches and strategies in response, such as slowing down the pace of the lesson or giving students hints for comprehension (Goleman, 1995). In IDLEs, the absence of teacher and/or the limitations of current technology may hinder the detection of confusion through digital interactions. Eventually, unresolved confusion is likely to contribute to learners abandoning the activity. Hence, a practical recommendation to developers and educators using IDLEs would be to elaborate techniques (e.g., based on observation of the behaviour or physiological responses) that allow real-time detection of students' emotions, and particularly confusion, during learning. The implementation of such techniques in learning environments would support interventions that attempt to tackle excessive levels of confusion or, possibly, to inversely induce some moderate confusion to keep students engaged in their learning activity.

For confusion detection, each measurement technique possesses its own strengths and limitations in terms of sensitivity, specificity, and time resolution, as well as cost and potential interference with the learning process. For this reason, a multimodal approach consisting of the integration of several indicators may be preferred to detect confusion (Hussain, Calvo, & Chen, 2014; Hussain, Monkaresi, &

Calvo, 2012). However, multimodal confusion detection requires multiple indicators of emotions from different sources such as behavioural observation, facial expressions, conversational cues, and physiological reactions (Pantic & Rothkrantz, 2003). In addition, an automatic processing of data from a combination of several measurement methods is likely to require high-level computing techniques. The training of relevant classifiers to integrate diverse signals can enhance the quality of confusion detection (Hussain, AlZoubi, Calvo, & D'Mello, 2011). Indeed, detecting confusion requires the development of classification schemes and fusion methods for the integration and the treatment of multimodal data, which is creating new challenges (Castellano, Gunes, Peters, & Schuller, 2014). For example, in a multimodal system involving data from different sources such as video, speech, and gestures, the detection of confusion can be done in different ways (Wagner, Andre, Lingenfelser, & Kim, 2011). The detection can indeed be performed either by combining together all data and using a single classifier (feature-level fusion), or on the other hand by using classifiers for each source of data and merging the ensemble of decisions into a single one (decision-level fusion). Of course each of the techniques has their own advantages and issues in regard of the treatment of temporary missing data, the efficiency of classification or the application in different environments (Wagner et al., 2011). However this effort seems worthwhile because it contributes to an increase in the power of detecting confusion since a fusion of multimodal indicators would outperform any form of detection based on single indicators (D'Mello & Kory, 2015).

The major limitation of most of the methods presented in this review is that the reliable detection of confusion based on behavioural and physiological indicators is still almost entirely limited to abstract and laboratory-based learning situations. Many of the methods reviewed in this paper may not be practical within real-world

IDLEs. However, the potential outcomes from experimentation using these methods are likely to significantly improve our knowledge of the dynamics of confusion during learning. Moreover, future studies of the detection of confusion in real-world learning situations will be useful for validating the efficacy of each indicator, for improving their individual detection capacity, and for assessing their viability as classifiers for developing predictive modelling of confusion.

Some techniques seem to be better candidates than others when considering their application by educators and developers in real-world learning settings. The ideal method would need to be technically available, be non-intrusive for learners (i.e., not imposing them to perform a additional competing task) and allow remote collection of data. The method that seems compliant with these constraints and that is already available would be the utilisation of the interaction data that participants generate when they are interacting with the interface of the learning system. Nevertheless, even if collecting activity log files from the system is technically simple, their interpretation as indicators of confusion requires further investigation. An outcome of an approach consisting of detecting confusion from several indicators during learning would be designing predictive models of the occurrence of confusion in IDLEs. The development of predictive models from log files is heavily reliant on having simultaneous external criterion measures of learners' confusion upon which to validate the models. For this reason, some techniques based on learning analytics seem to be promising because of their ability to be derived in laboratory settings and implemented in ecological environments. Using this approach ultimately allows the use of data from students' interactions within IDLEs to identify specific patterns linked to emotional states during learning, including confusion, without the need of any additional sensors (See Baker et al., 2012; Pedro, Baker, Gowda, & Heffernan, 2013). These are novel approaches that reflect advances in the research areas of learning analytics, educational data mining, machine learning, and more generally affective computing.

The algorithms that will make possible a fast and accurate detection of learners' confusion in their interactions with a IDLE could be developed and validated with the recording of physiological indicators of emotions. Even if most of the physiological measures presented in this article cannot be easily implemented in real-world learning environments, their potential valuable contribution to the challenging project of developing valid and reliable indicators for the detection and prediction of learners' confusion should not be underestimated. In the future, it is conceivable that some predictive models will not only be able to detect the absence and presence of confusion, but also predict the extent to which experienced confusion will be subsequently productive in terms of learning outcomes. If confusion could be reliably detected, adaptive interventions could be designed to support learners navigate within their zone of optimal confusion. Like a teacher in the classroom, IDLEs could detect when students are getting confused, bored or disengaged and produce an adaptive, tailored response to meet the individual needs of each learner at specific stages of the learning experience.

Acknowledgments

A Special Research Initiative of the Australian Research Council supported this research: ARC-SRI Science of Learning Research Centre (project number SRI20300015).

References

Aasman, J., Mulder, G., & Mulder, L. J. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *29*(2), 161–170.

Allen, J. J., Chambers, A. S., & Towers, D. N. (2007). The many metrics of cardiac chronotropy: A pragmatic primer and a brief comparison of metrics. *Biological Psychology*, *74*(2), 243–262.

Allen, V., MacCann, C., & Matthews, G. (2014). Emotional Intelligence in Education: from Pop to Emerging Science. In *International handbook of emotions in education* (p. 162).

Apps, M. A., Lesage, E., & Ramnani, N. (2015). Vicarious Reinforcement Learning Signals When Instructing Others. *The Journal of Neuroscience*, *35*(7), 2904–2913.

Arguel, A., & Lane, R. (2015). Fostering deep understanding in geography by inducing and managing confusion: an online learning approach. In *Globally connected, digitally enabled. Proceedings ascilite 2015* (pp. 22–26). Perth.

Asteriadis, S., Tzouveli, P., Karpouzis, K., & Kollias, S. (2009). Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment. *Multimedia Tools and Applications*, *41*(3), 469–493.

Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241. doi:10.1016/J.Ijhcs.2009.12.003

Baker, R. S. J. D., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., … Rossi, L. (2012). Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. *International Educational Data Mining Society*.

Bradley, M. M., & Lang, P. J. (2000). Emotion and motivation. *Handbook of Psychophysiology*, *2*, 602–642.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281.

Caballe, S., Barolli, L., Feidakis, M., Matsuo, K., Xhafa, F., Daradoumis, T., & Oda, T. (2014). A Study of Using SmartBox to Embed Emotion Awareness through Stimulation into E-learning Environments. In *2014 International Conference on Intelligent Networking and Collaborative Systems (INCoS)* (pp. 469–474). doi:10.1109/INCoS.2014.9

Calvo, R. A., & D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37. doi:10.1109/T-AFFC.2010.1

Calvo, R. A., D'Mello, S., Gratch, J., & Kappas, A. (2014). *The Oxford Handbook of Affective Computing*. Oxford University Press.

Castellano, G., Gunes, H., Peters, C., & Schuller, B. (2014). Multimodal Affect Recognition for Naturalistic Human-Computer and Human-Robot Interactions. In *The Oxford Handbook of Affective Computing* (p. 246).

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. *The Handbook of Emotion Elicitation and Assessment*, 203–221.

Conati, C., Hoque, E., Toker, D., & Steichen, B. (2013). When to Adapt: Detecting User's Confusion During Visualization Processing. In *UMAP Workshops*.

Craig, S., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive–affective states during learning. *Cognition and Emotion*, *22*(5), 777–788.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004a). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, *29*(3), 241–250.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004b). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, *29*(3), 241–250.

Darwin, C., Ekman, P., & Prodger, P. (1998). *The Expression of the Emotions in Man and Animals*. Oxford University Press.

DeLucia, P. R., Preddy, D., Derby, P., Tharanathan, A., & Putrevu, S. (2014). Eye Movement Behavior During Confusion Toward a Method (Vol. 58, pp. 1300–1304). Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, SAGE Publications.

Dimberg, U. (1990). For Distinguished Early Career Contribution to Psychophysiology: Award Address, 1988. *Psychophysiology*, *27*(5), 481–494. doi:10.1111/j.1469-8986.1990.tb01962.x

D'Mello, S., Craig, S. D., Witherspoon, A., Mcdaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, *18*(1-2), 45–80.

D'Mello, S., & Graesser, A. (2006). Affect Detection from Human-Computer Dialogue with an Intelligent Tutoring System. In J. Gratch, M. Young, R. Aylett, D. Ballin, & P. Olivier (Eds.), *Intelligent Virtual Agents* (pp. 54–67). Springer Berlin Heidelberg.

D'Mello, S., & Graesser, A. (2007). Mind and body: Dialogue and posture for affect detection in learning environments. *Frontiers in Artificial Intelligence and Applications*, *158*, 161.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. doi:10.1016/J.Learninstruc.2011.10.001

D'Mello, S., & Graesser, A. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychologica*, *151*, 106–116. doi:10.1016/J.Actpsy.2014.06.005

D'Mello, S., & Kory, J. (2015). A Review and Meta-Analysis of Multimodal Affect Detection Systems. *ACM Computing Surveys (CSUR)*, *47*(3), 43.

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

Durso, F. T., Geldbach, K. M., & Corballis, P. (2012). Detecting Confusion Using Facial Electromyography. *Human Factors*, *54*(1), 60–69. doi:10.1177/0018720811428450

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. *Handbook of Applied Cognition*, 283–314.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169–200.

Elkins, A., Zafeiriou, S., Pantic, M., & Burgoon, J. (2015). Unobtrusive deception detection.

Feidakis, M., Daradoumis, T., Caballé, S., & Conesa, J. (2013). Measuring the Impact of Emotion Awareness on e-learning Situations. In *2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)* (pp. 391–396). doi:10.1109/CISIS.2013.71

Feidakis, M., Daradoumis, T., Caballé, S., Conesa, J., Feidakis, M., Daradoumis, T., … Conesa, J. (2014). Embedding emotion awareness into e-learning environments. *International Journal of Emerging Technologies in Learning (iJET)*, *9*(7), 39–46.

Fowles, D. C. (1980). The three arousal model: Implications of Gray's two‐factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology*, *17*(2), 87–104.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Graesser, A. C. (2011). Learning, thinking, and emoting with discourse technologies. *American Psychologist*, *66*(8), 746–757. doi:10.1037/a0024974

Graesser, A. C., & D'Mello, S. (2012). Emotions during the Learning of Difficult Material. *Psychology of Learning and Motivation, Vol 57*, *57*, 183–225. doi:10.1016/B978-0-12-394293-7.00005-4

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory & Cognition*, *33*(7), 1235–1247. doi:10.3758/BF03193225

Grafsgaard, J. F., Boyer, K. E., & Lester, J. C. (2011). Predicting facial indicators of confusion with hidden Markov models. In *Affective Computing and Intelligent Interaction* (pp. 97–106). Springer.

Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2013). Automatically Recognizing Facial Expression: Predicting Engagement and Frustration. In *Proceedings of the 6th International Conference on Educational Data Mining*.

Healey, J. (2014). Physiological Sensing of Emotion. In *The Oxford Handbook of Affective Computing* (p. 204).

Hester, R., Barre, N., Murphy, K., Silk, T. J., & Mattingley, J. B. (2008). Human Medial Frontal Cortex Activity Predicts Learning from Errors. *Cerebral Cortex*, *18*(8), 1933–1940. doi:10.1093/cercor/bhm219

Huang, C.-N., Chen, C.-H., & Chung, H.-Y. (2004). The review of applications and measurements in facial electromyography. *Journal of Medical and Biological Engineering*, *25*(1), 15–20.

Hussain, M. S., AlZoubi, O., Calvo, R. A., & D'Mello, S. K. (2011). Affect detection from multichannel physiology during learning sessions with AutoTutor. In *Artificial Intelligence in Education* (pp. 131–138). Springer.

Hussain, M. S., Calvo, R. A., & Chen, F. (2014). Automatic cognitive load detection from face, physiology, task performance and fusion during affective interference. *Interacting with Computers*, *26*(3), 256–268.

Hussain, M. S., Monkaresi, H., & Calvo, R. A. (2012). Combining Classifiers in Multimodal Affect Detection. Presented at the Tenth Australasian Data Mining Conference (AusDM2012), Sydney.

Kapoor, A., Mota, S., & Picard, R. W. (2001). Towards a learning companion that recognizes affect. In *AAAI Fall symposium* (pp. 2–4).

Kemp, A. H., Krygier, J., & Harmon-Jones, E. (2014). Neuroscientific perspectives of emotions. In *The Oxford Handbook of Affective Computing*.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776–785. doi:10.1111/1469-8986.00078

Lehman, B., D'Mello, S., & Graesser, A. (2012). Confusion and complex learning during interactions with computer learning environments. *Internet and Higher Education*, *15*(3), 184–194. doi:10.1016/J.Iheduc.2012.01.002

Lehman, B., D'Mello, S., & Graesser, A. (2013). Who Benefits from Confusion Induction during Learning? An Individual Differences Cluster Analysis (pp. 51–60). Presented at the Artificial Intelligence in Education, Springer.

Lehman, B., & Graesser, A. (2015). To Resolve or not to Resolve? that is the Big Question About Confusion. In *Artificial Intelligence in Education* (pp. 216–225). Springer.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363–384.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, *11*(4), 357–380.

Lipson, A. (1992). The confused student in introductory science. *College Teaching*, *40*(3), 91–95.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* (pp. 298–305). IEEE.

Lodge, J. M., & Kennedy, G. (2015). Prior knowledge, confidence and understanding in interactive tutorials and simulations. In *Globally connected, digitally enabled. Proceedings ascilite 2015* (pp. 190–201). Perth.

Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied*, *2*(2), 164–177. doi:10.1037/1076-898X.2.2.164

Macdowell, K. A., & Mandler, G. (1989). Constructions of Emotion - Discrepancy, Arousal, and Mood. *Motivation and Emotion*, *13*(2), 105–124. doi:10.1007/Bf00992957

Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. WW Norton.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies*, *65*(4), 329–347.

Marosi, E., Bazán, O., Yañez, G., Bernal, J., Fernández, T., Rodríguez, M., … Reyes, A. (2002). Narrow-band spectral measurements of EEG during emotional tasks. *International Journal of Neuroscience*, *112*(7), 871–891.

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. (1990). Eye movement indices of mental workload. *Acta Psychologica*, *75*(1), 75–89.

Narciss, S. (2004). The Impact of Informative Tutoring Feedback and Self-Efficacy on Motivation and Achievement in Concept Learning VL -.51. *Experimental Psychology*, (3), pp. doi:10.1027/1618-3169.51.3.214

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71.

Paas, F., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1), 419–430.

Pachman, M., Arguel, A., Lockyer, L., Kennedy, G., & Lodge, J. M. (in press). Eye tracking and early detection of confusion in digital learning environments: proof of concept. *Australasian Journal of Educational Technology*.

Pantic, M., & Rothkrantz, L. J. M. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, *91*(9), 1370–1390. doi:10.1109/JPROC.2003.817122

Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes (pp. 117–124). Presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM.

Pecchinenda, A., & Smith, C. A. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition & Emotion*, *10*(5), 481–504.

Pedro, M. O. Z. S., Baker, R. S. J. d, Gowda, S. M., & Heffernan, N. T. (2013). Towards an Understanding of Affect and Knowledge from Student Interaction with an Intelligent Tutoring System. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 41–50). Springer Berlin Heidelberg.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice VL -.18. *Educational Psychology Review*, (4), pp. doi:10.1007/s10648-006-9029-9

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). Washington, DC: American Psychological Association; US.

Postma-Nilsenová, M., Postma, E., & Tates, K. (2015). Automatic Detection of Confusion in Elderly Users of a Web-Based Health Instruction Video. *Telemedicine and E-Health*.

Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Frontiers in Neuroscience*, *5*. doi:10.3389/fnins.2011.00115

Rodrigo, M. M. T., Baker, R. S., Jadud, M. C., Amarra, A. C. M., Dy, T., Espejo-Lahoz, M. B. V., … Tabanao, E. S. (2009). Affective and behavioral predictors of novice programmer achievement. In *ACM SIGCSE Bulletin* (Vol. 41, pp. 156–160). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1562929

Rozin, P., & Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, *3*(1), 68–75. doi:10.1037/1528-3542.3.1.68

Sato, W., Fujimura, T., & Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *International Journal of Psychophysiology*, *70*(1), 70–74.

Shan, C., & Braspenning, R. (2010). Recognizing Facial Expressions Automatically from Video. In H. Nakashima, H. Aghajan, & J. C. Augusto (Eds.), *Handbook of Ambient Intelligence and Smart Environments* (pp. 479–509). Springer US.

Shen, L., Wang, M., & Shen, R. (2009). Affective e-learning: Using 'emotional' data to improve learning in pervasive learning environment. *Journal of Educational Technology & Society*, *12*(2), 176–189.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 0002764213498851.

Soliman, M., & Guetl, C. (2010). Intelligent pedagogical agents in immersive virtual learning environments: A review. In *MIPRO, 2010 proceedings of the 33rd international convention* (pp. 827–832). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5533534

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

Tan, H. Z., Lu, I., & Pentland, A. (1997). The chair as a novel haptic user interface. In *Proc. of the Workshop on Perceptual User Interfaces* (pp. 19–21).

Tattersall, A. J., & Hockey, G. R. J. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(4), 682–698.

Umemuro, H., & Yamashita, J. (2003). Detection of user's confusion and surprise based on pupil dilation. *The Japanese Journal of Ergonomics*, *39*(4), 153–161. doi:10.5100/jje.39.153

van Dooren, M., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, *106*(2), 298–304.

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.

Wagner, J., Andre, E., Lingenfelser, F., & Kim, J. (2011). Exploring fusion methods for multimodal emotion recognition with missing data. *Affective Computing, IEEE Transactions on*, *2*(4), 206–218.

Wang, H., Chignell, M., & Ishizuka, M. (2006). Empathic tutoring software agents using real-time eye tracking. In *Proceedings of the 2006 symposium on Eye tracking research & applications* (pp. 73–78). ACM.

Wang, H., Li, Y., Hu, X., Yang, Y., Meng, Z., & Chang, K. (2013). Using EEG to Improve Massive Open Online Courses Feedback Interaction. In *AIED Workshops*.

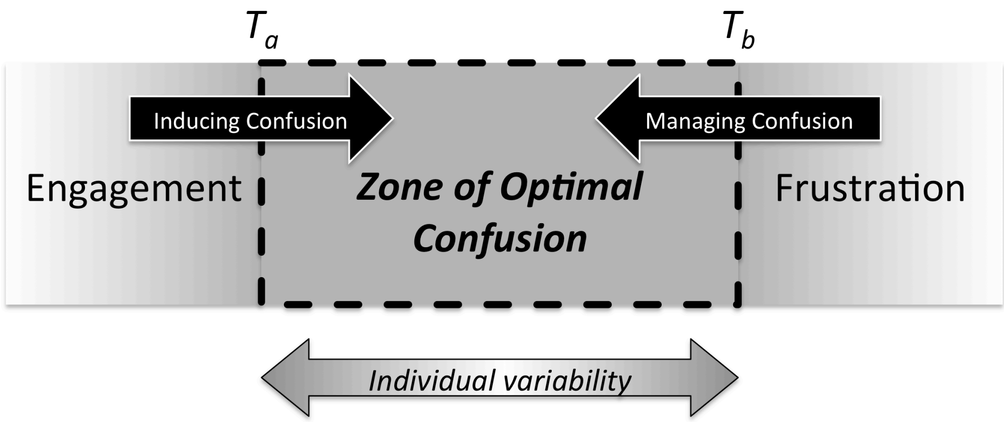

Figure 1: A representation of the *zone of optimal confusion* and possible external interventions (Arguel & Lane, 2015), reprinted with permission.