International Journal of Child-Computer Interaction 34 (2022) 100536

Contents lists available at ScienceDirect

International Journal of Child-Computer Interaction

journal homepage: www.elsevier.com/locate/ijcci

Research paper

Children's facial expressions during collaborative coding: Objective versus subjective performances

Kshitij Sharma *, Sofia Papavlasopoulou, Michail Giannakos

Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

Article history:
Received 7 January 2022
Received in revised form 14 July 2022
Accepted 27 August 2022
Available online 12 September 2022

Keywords:
Control value theory
Facial expressions
Emotions
Affect
Collaborative learning
Collaborative coding

ABSTRACT

Educational research has used the information extracted from facial expressions to explain learning performance in various educational settings like collaborative learning. Leveraging this, we extracted the emotions based upon two different theoretical frameworks from videos with children aged 13–16 while collaborating to create games using Scratch. The two sets of emotions are based on the control value theory (happiness, sadness, anger, surprise) and the education-specific expressions (frustration, boredom, confusion, delight). We computed the groups' objective performance, which was calculated based on their created artifacts. We divided them into high and low performance and compared them based on individual emotions' duration and the transitions among the emotions. We also used the subjective indication of their perceived performance from a self-reported questionnaire, divided them into another performance category, and did a similar analysis with the objective performance. Results show that the objective performance is better explained by the education-specific emotions and the negative valance emotions from the control value theory-based emotions. On the other hand, subjective performance is better explained by the control value theory based on emotions. Based on the results, we suggest implications both for the instructors and students.

1. Introduction

Computational thinking and coding activities for children in K-12 education have been growing over the past years. Many European countries have already incorporated computer science (CS) into the school curricula for some years (Brown, Sentance, Crick, & Humphreys, 2014) while several others followed (Bocconi et al., 2016; Tuhkala, Wagner, Nielsen, Iversen, & Kärkkäinen, 2018). Organizations like Computer Science Teachers Association (CSTA), Informatics Europe, and the Cyber Innovation Center, to mention a few, support and encourage CS education with practices, while others, like "Code.org" and "code the future" are offering many resources to support coding. In addition, the existence of low-cost mini-computers such as Micro: Bit and Raspberry Pi, together with educational programming languages like Scratch, Alice, and Blockly, have contributed to a large-scale adoption from children. More and more coding activities appear in both in- and out-of-school settings, during which children have the opportunity to develop digital skills, turn into creative developers of their projects and gain confidence at different levels following technology that transforms our digital society.

Many times, CS and coding activities are based on Papert's constructionism (Papert, 1990) that emphasizes the importance of how the process of creating a shared and meaningful artifact is the key to gaining knowledge. Available educational child-friendly tools and practices (e.g., Scratch, K-12 CS framework) are good examples that can offer fruitful learning experiences to children allowing them to learn how to code. These tools and practices can also enhance their computational thinking, problem-solving and collaborative skills, Denner, Werner, and Ortiz (2012), Papavlasopoulou, Giannakos, and Jaccheri (2017).

Collaborative making and coding activities for children are valued for their engagement and building of knowledge and for enhancing the social setting. During these activities, children share their experiences and together interact having a common purpose; they overcome possible difficulties they meet in the process and share their emotions related to the individual, the interaction with technology, and the group dynamics (Denner, Werner, Campe, & Ortiz, 2014; Israel et al., 2016; Stahl, Law, Cress, & Ludvigsen, 2014). In technology-based settings, emotions are essential drivers of learning and can be shaped by the different aspects of the settings and learners' experience (Loderer, Pekrun, & Lester, 2018). Kort et al. 2001 (Kort, Reilly, & Picard, 2001) listed the emotions involved in learning and proposed a model with the phases of learning in emotions that cycle from positive to negative. For example, a student may start dealing

* Corresponding author.
E-mail addresses: kshitij.sharma@ntnu.no (K. Sharma), spapav@ntnu.no (S. Papavlasopoulou), michailg@ntnu.no (M. Giannakos).

https://doi.org/10.1016/j.ijcci.2022.100536

with a task with confusion, possibly due to difficulty, followed by frustration, and then by happiness related to success (Kinnunen & Simon, 2010, 2012). Children's individual and collaborative emotional behaviors during activities like coding must be studied closely to understand the learning experience and design systems and activities to support them efficiently. Compared to performance and learning gains, learners' emotions are harder to be measured (Picard et al., 2004). Most of the time, affect in learning contexts and specifically in coding activities have been examined through qualitative and quantitative measurements like questionnaires, observations, and interviews (D'Mello, 2013; Jordan & McDaniel Jr, 2014; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011; Zhang, Markopoulos, Bekker, Schüll, & Paule-Ruíz, 2019). However, some efforts suggest computing surface-level affect behaviors based on gaze, facial expressions, head movements, and gestures (Kapoor et al., 2001), extraction of joint emotional states from videos (Sharma et al., 2019). These surface-level affect behaviors help develop an instrument that focuses "at the moment" individual and collaborative interactions during computing collaborations (Israel et al., 2016).

Computer science, particularly programming, has historically been complex for novices to learn (Jenkins, 2002). Collaborative programming has been proposed as a way to address novice programmers' challenges by supporting them in collaborating with another student (McDowell, Werner, Bullock, & Fernald, 2003; Williams & Upchurch, 2001). Using pair programming to address these challenges is not a new idea, and significant research has shown it can benefit students. Specifically, in terms of learning, research shows that when students engage in collaborative programming, they can solve problems more quickly and effectively than those working individually (Hannay, Dybå, Arisholm, & Sjøberg, 2009; McDowell, Werner, Bullock, & Fernald, 2002; Radermacher & Walia, 2011; Williams & Upchurch, 2001). In addition to the impact on learning, collaborative programming can also positively impact students' interest and attitudes towards computer science (Hannay et al., 2009). Their confidence in their work (McDowell et al., 2003). The research found that students are more satisfied when working in groups/pairs than when working individually (Salleh, Mendes, & Grundy, 2010). In terms of drop-outs, students that engage in collaborative programming have higher persistence in computer science courses, and increased retention (O'Donnell, Buckley, Mahdi, Nelson, & English, 2015; Porter & Simon, 2013; Umapathy & Ritzhaupt, 2017).

These positive impacts of collaborative programming occur when students implement the processes correctly (Umapathy & Ritzhaupt, 2017). Not all such programming interactions are successful (Werner & Lester, 2001), and a non-successful group can lead to students being less productive or failing to complete the assignment (Werner, Hanks, & McDowell, 2004). If students are poorly supported, it may impact the students' learning (Radermacher & Walia, 2011) and their attitudes towards computer science (Seyam & McCrickard, 2016). Moreover, the negative experiences can discourage students from working collaboratively in the future (Schultz, Wilson, & Hess, 2010). Students need support and guidance to engage in beneficial collaborative programming processes (Govender & Govender, 2014). One of the ways to provide such support is by tracing their affective states while collaboratively coding, as we will see in the other sections of this contribution.

Learning, and by extension, collaborative learning is a multifaceted phenomenon. This study is a step toward better understanding how individual emotional state changes over time during children's learning experience and interaction with coding tasks and team interactions. Moreover, emotions and affect are related to the both short-term (Bowden, 2015; Cheng, Huang, & Hsu, 2020) and long-term (Meyer & Turner, 2002; Shen, Wang, & Shen, 2009) learning outcomes. Apart from an objective evaluation of performance that can indicate a learning gain, it is important to consider children's perceived performance. Perceived performance is important, as their satisfaction from a positive experience and perception influences their intrinsic motivation, self-efficacy, and engagement. Therefore, it is important for education researchers to consider the diversity of emotional/affective states in the academic settings by considering a diverse set of emotions experienced by students at school (Pekrun, Goetz, Titz, & Perry, 2002a).

This study explores how children (age 13–16 years old) experience the process of collaboratively learning how to code while creating a shared artifact (i.e., a game). These children were participating in our making-based coding activity. Therefore, we propose a quantitative analysis to capture the children's emotional state during the activity using video recordings from a webcam. In addition, we collected and analyzed the created artifacts as an indication of their task-based performance (high and low). Along with the task-based performance, which is the objective performance, we also include a subjective performance measurement from children's answers to a questionnaire regarding the perceived learning performance of their team. We analyzed the data collected from two sets of emotional states, that is, Control Value Theory (CVT, Pekrun (2006)) and the Affective Framework for Learning (AFL, D'Mello and Graesser (2012)). We explore the relation between these two sets of emotional states and the objective and subjective measurements of performance in the collaborative coding sessions. To the best of our knowledge, this is the first attempt to include two performance metrics and associate them with two sets of affective states/emotions based on two different theoretical frameworks. By doing so, we aim to bring forth a set of guidelines that can influence the development of scaffolding tools in various contexts.

In this paper, we have used two different theories about students' emotions to inform our analysis. This practice is not shown in the other related avenues. The main reason for incorporating the two theories about the students' emotions is that one is about the overall achievement, and the other explains the process of evolution. CVT provides a detailed overview of the factors related to the optimal academic development, of the concerned students, over a relatively long period (Putwain, Becker, Symes, & Pekrun, 2018; Putwain, Pekrun, et al., 2018). On the other hand, the education-specific emotions are more dynamic and therefore require to be monitored over a relatively shorter period (Mitra & Chavan, 2019; Simonton & Garn, 2019). Pekrun's theory (CVT) has been used to show the relation between positive and negative emotions and the overall achievement (Pekrun, Frenzel, Goetz, & Perry, 2007), and the dynamics of students' emotion has been explained using the cognitive disequilibrium theory (CDT) (D'Mello & Graesser, 2010). As the name, cognitive "disequilibrium" itself suggests that the framework of education-specific emotions is related to the dynamic nature of the emotions and the CVT is concerned with the overall extent of the achievement emotions. The two performance measurements used in this paper, i.e., the perceived and objective performance, share similar differences as the two driving theoretical foundations. In the present study, the objective performance is measured using the code produced at the end of the coding activity. In contrast, the subjective performance was measured by overall perception of children about their learning during the coding activity. The objective performance resulted from dynamic, collaborative interactions and learning processes. On the other hand, the objective performance resulted from the self-reported overall perception of their performance. To be able to capture the overall emotions and the process-based emotions, we included both the CVT-based and education-specific emotions while studying the relationship between the children's emotions and their performance in the collaborative coding performances (both objective and subjective).

Specifically, in this paper, we address the following research questions:

RQ1: How does the **education** specific emotions (i.e., boredom, frustration, confusion, delight, neutral) and their evolution relate to the **objective** performance levels (high/low) in collaborative coding sessions?

RQ2: How does the **CVT** specific emotions (i.e., happiness, sadness, anger, surprise, neutral) and their evolution relate to the **objective** performance levels (high/low) in collaborative coding sessions?

RQ3: How does the **education** specific emotions (i.e., boredom, frustration, confusion, delight, neutral) and their evolution relate to the **subjective** performance levels (high/low) in collaborative coding sessions?

RQ4: How does the **CVT** specific emotions (i.e., happiness, sadness, anger, surprise, neutral) and their evolution relate to the **subjective** performance levels(high/low) in collaborative coding sessions?

Concretely, we contribute the following through this paper.

- We provide empirical evidence comparing the two most used theoretical frameworks for emotions in explaining the difference between various performance levels (i.e., high performance versus low performance) in the context of collaborative coding.
- We provide empirical evidence comparing the two most used theoretical frameworks for emotions in explaining the difference between various performance measurements (i.e., subjective performance versus objective performance) in the context of collaborative coding.
- We provide insights into the design of scaffolding tools for students and instructors.

The remainder of the paper is organized as follows. The second section provides the Theoretical frameworks describing the control value theory and education-specific emotions, as well as the background on the collaborative learning and coding activities with children. The third section presents the details of the data collection, variables involved, and the analysis methods. The fifth section discusses the results and their implications. Finally, the sixth section concludes the paper.

2. Related work

During K-12 CS/CT activities, children often have the chance to not only work individually on tasks but, depending on the design of the activity, engage in a collaborative coding experience with peers. Although it is not an easy process, collaboration is an important part of learning CS and coding (Tsan, Lynch, & Boyer, 2018). Building on Vygotsky (Vygotsky, 1978), and Dewey (Dewey, 2018), it is shown that through collaboration, children construct meaning and knowledge. A common purpose allows children to learn from each other, share responsibilities, and confront difficulties. Compared to working alone, when children work in teams can be engaged in discussions relevant to the completion of the task, be aware of their learning, be persistent in challenging tasks and confront struggles (Goos, Galbraith, & Renshaw, 2002; Hmelo-Silver & Barrows, 2008; Werner & Denning, 2009). Roschelle and colleagues refer to the notion of joint problem space (JpS), which is essential for collaborative learning as it includes the shared conception, goals, and knowledge (Roschelle, 1992; Teasley & Roschelle, 1993). As children collaborate to find a solution to a problem, their metacognitive thinking is also uncovered (Kuhn, 2015). Children's thinking process is shown from their interactions and negotiations; the way children approach these actions will result in collaboration.

The development of computational artifacts is not simple or linear. On the contrary, it is an iterative process of decisions, trials, and testing (Brennan & Resnick, 2012). Studies show the benefits of collaborative learning for children's performance and cognition (Barron, 2000; Chan, 2013). While creating and debugging a game, girls who had an effective collaboration were trying more on their own before asking for help from the instructors (Denner, 2007). In their study, Jordan and McDaniel focused on how 5th-grade students influenced each other during collaboration in a robotics engineering activity. While working on problem-solving, students experienced content but also the uncertainty that was either directly resolved or followed by supportive or unsupportive ways of action for the peers (Jordan & McDaniel Jr, 2014). Denner et al. showed that middle school children with low prior computer use who worked in pairs using the Alice programming environment increased their programming knowledge (Denner et al., 2014). Their study suggests that when one of the partners has more experience, the other can still learn. Sullivan and Wilson (2015) suggested playful talk as a way to avoid conflicts and competitive attitude of students working in coding and other physics/robotics' curriculum tasks (Sullivan & Wilson, 2015). In this way, the tensions are decreased, and opportunities to learn are opened for low-status group members.

2.1. Emotions in education and CCI

Children's emotions and their affective states are and have been a major direction of CCI research, with several studies evaluating and/or exploring performance (Sharma et al., 2019), enjoyment (Leite et al., 2009), usability (Giannakos, Chorianopoulos, Inkpen, Du, & Johns, 2013; Tsai, Lo, & Chen, 2012), engagement (Leite, Henriques, Martinho, & Paiva, 2013) and learning processes (Sridhar, Chan, & Nanayakkara, 2018). Moreover, emotions/expressions/affective states have been used in educational research to improve students' interaction (Harrold, Tan, Rosser, & Leong, 2014; Perry & Aragon, 2012; Suzuki, 2015), provide feedback (Spaulding, Gordon, & Breazeal, 2016; Tsai et al., 2012; Zhang, 2008), and evaluate/understand task-based performance (Jiménez et al., 2018). A systematic literature review has been conducted about affective states and emotions in educational settings (Reis et al., 2018). The results show that most of the selected studies emphasized increasing emotional awareness among students during collaborative sessions and the usefulness of emotions for orchestrating their interactions and better group formation (Reis et al., 2018). Among other reviews (Loderer, Pekrun, & Lester, 2020; Reis et al., 2018; Zhang, Markopoulos, & Bekker, 2020) about using CVT-based emotions as a process analytics tool, the results show a positive correlation between enjoyment and achievement and a negative correlation between anxiety and learning strategy. The authors also concluded from the review that it is becoming paramount in technology-based learning to understand and support emotional processes (Loderer et al., 2020). The results of another literature review paper (Zhang et al., 2020) show that boredom, satisfaction, relaxation, interest, and curiosity were among the most studied emotions regarding design-based learning. However, most of the reviewed papers used non-automatic means to collect students' emotions during learning sessions (e.g., survey, video coding, interviews, observation, questionnaire) (Zhang et al., 2020).

Emotions in CCI research have been measured through multiple data collection modes such as facial features (Kapoor & Picard, 2005; Leite et al., 2009; Tsai et al., 2012), physiological data (Delaborde, Tahon, Barras, & Devillers, 2009; Leite et al., 2013; Sridhar et al., 2018; Takano & Suzuki, 2014), and self reports (Giannakos et al., 2013; Shahid, Krahmer, Swerts, & Mubin, 2010). All the techniques mentioned above have their respective advantages and disadvantages. For example, questionnaire interviews are considered to be favorable. However, because they are done at a specific frequency and multiple intervals during the task performance, they are susceptible to distraction in learning contexts. Moreover, identical timing and frequency might not work for tasks with different requirements and/or complexity. Such methods of measuring emotions cannot account for rapid changes in the learner's facial expressions. Experimenter observations of facial expressions can be a tedious task. They require extensive training and an extensive amount of time. Even then, the observations could not capture the rapidly changing nature of emotions in a collaborative task. Finally, the increasing availability of the off-the-shelf-sensors supports the automated measurement of emotions, even when no apparent change in task performance can be detected. However, some of the modern off-the-shelf sensors (e.g., EEG, fMRI, wristbands) might be discomforting for the users over a long period of interaction. For example, fMRI machines limit the motion and the interaction with the learning technology. With the facial recording, there is one advantage that the cameras are, in most cases, unintrusive. However, recording students' faces, especially children, might raise certain ethical and privacy-related issues (for details, see Sharma and Giannakos (2021)). In a review about affective computing in the educational setting, the results show that the increasing use of off-the-shelf sensor technology has enabled it to automatically capture the affective states in various learning settings. Moreover, emotions extracted through facial videos were the second most used automatic method to capture affective states (only the skin conductance was used more often) (Wu, Huang, & Hwang, 2016).

There are two primary strands when it comes to utilizing the emotions that are considered important in educational settings. The first set is derived from the Control Value Theory (e.g., sadness, happiness, surprise, anger, disgust) (Pekrun, 2006) and the second set comprises the affective states (e.g., confusion, frustration, boredom, delight) (D'Mello & Graesser, 2012). In a review about affective states in various learning scenarios, it was found that both the CVT-based emotions and education-specific emotions were used by the top 50 percentile of the contributions (Wu et al., 2016). Following are the brief descriptions of the two major theoretical frameworks.

Control Value Theory (CVT) is primarily concerned with the emotions that could be directly linked to the achievement outcomes (Pekrun et al., 2007, 2002a; Pekrun, Goetz, Titz, & Perry, 2002b). CVT explains how and why the students' emotions contribute to academic and nonacademic outcomes (Pekrun et al., 2007; Simonton & Garn, 2020). It provides a comprehensive framework for investigating the relationship between students' emotional antecedents and performance-based outcomes (Mitra & Chavan, 2019). There are three dimensions in the taxonomy that describes the achievement emotions: object focus (Activity versus Outcome, whether to focus on the process or the outcome), activation (Activating versus Deactivating), and valance (Positive versus Negative). The captured emotion depends on the perceived control and value appraisals. If the student feels that they have control over the learning activity and they value the activity positively, we observe positive emotions (e.g., happiness, contentment, relief); otherwise, we observe negative emotions (e.g., anger, sadness) (Pekrun et al., 2007; Tze, Parker, & Sukovieff, 2022). The perceived value, of the learning activity, on the other hand, can either be intrinsic (process-based) or extrinsic (outcome-based) (Pekrun et al., 2007; Tze et al., 2022).

The education specific emotions are based upon the theoretical framework known as the cognitive disequilibrium theory (CDT) (Barrouillet, 2015; Berlyne, 1960). This theory connects the affective and cognitive dimensions of learning (D'Mello & Graesser, 2010). The core idea of the theory is that cognitive disequilibrium is important for learning and comprehension (D'Mello & Graesser, 2010, 2012). It is explained in terms of state transition. That is, the equilibrium state corresponds to the Flow state as presented by Csikszentmihalhi (Csikszentmihalhi, 2020). As soon as an impasse is detected between the problem at hand and the students' processes, it instigates confusion that, in turn, creates disequilibrium. If the impasse is resolved, then the students go back to equilibrium, and if momentary failure to resolve the impasse turns to frustration (feeling stuck) and the subsequent long-term failure results in boredom (disengagement).

Previous research, with the control value theoretic emotions, has shown that happiness is correlated with success (Fredrickson, 1998) and anger is correlated with failure (Bless, 2000). According to control value theory, (Pekrun, 2006), happiness is related to high prospective success, anger is related to retrospective failure, and sadness is related to the high negative activity. Emotions were also related to the competence belief and the value students attribute to a particular domain (Frenzel, Pekrun, & Goetz, 2007). Emotion is an essential part of studying motivation in classroom interactions since teachers' instructional and interpersonal responses to students are often governed by emotions (Meyer & Turner, 2002).

There is little support for a direct relationship between emotions and learning performance (Linnenbrink & Pintrich, 2002, 2003); however, frustration is a common feeling among students involved in online collaborative learning experiences (Capdeferro & Romero, 2012). Studies have reported on the relationship between gender, performance, and emotional showcase. For example, high-performing girls show less positive emotions than high-performing boys (Seegers & Boekaerts, 1993). Another facet of studies about emotions in educational contexts shows how emotions influence how information is processed. Happiness/joy results in novel and creative actions (Fredrickson, 1998), positive emotions also promote the engagement in metacognitive processing (Linnenbrink & Pintrich, 2003) which is beneficial for long-term learning. On the other hand, negative emotions focus on environmental specific details (Bless, 2000) also, negative emotions lead to a lack of elaboration (Pekmn et al., 2002). Moreover, negative affect was associated with lower learning goals (Meyer & Turner, 2002); while positive affect was associated with the interest in a given topic (Ainley, Corrigan, & Richardson, 2005).

Based on the education-specific facial expressions, frustration was a common feeling among students involved in online collaborative learning (Makewa, Gitonga, Ngussa, Njoroge, & Kuboja, 2014). In contrast, boredom and confusion are related to poor academic performance (Baker, D'Mello, Rodrigo, & Graesser, 2010; Dowd, Araujo, & Mazur, 2015). Frustration is mostly associated when students are interacting/engaging with learning material/context that is complex (Di Leo, Muis, Singh, & Psaradellis, 2019; Liu, Pataranutaporn, Ocumpaugh, & Baker, 2013). It is often seen as the result of annoyance and fear of failure (Ford & Parnin, 2015; Harrington, 2005). Multiple studies have found mixed results while investigating the relationship between frustration and learning outcomes. For example, McQuiggan and Lester found a negative relationship between the learning outcomes and frustration (McQuiggan & Lester, 2007); Pardos et al. showed positive correlations (Pardos, Baker, San Pedro, Gowda, & Gowda, 2014). On the other hand, D'Mello et al. and Rodrigo et al. did not find a significant relationship between learning outcomes and frustration (D'Mello, 2013; Rodrigo et al., 2012). At the same time, Liu et al. reported mixed outcomes from multiple studies (Liu et al., 2013).

Regarding confusion, Richey and colleagues (Richey et al., 2019) found that the induced confusion could be beneficial for learning. In some other studies, the combination of frustration and confusion had been studied in various learning scenarios. Confusion was shown to be positively correlated in some studies (D'Mello, Lehman, Pekrun, & Graesser, 2014; Lehman et al., 2013) while in others. Confusion was negatively correlated with the learning outcome (Rodrigo et al., 2009; Schneider et al., 2016). Combining confusion with frustration, Liu and colleagues found that frustration following a long confusion duration was negatively correlated to the learning outcome (Liu et al., 2013). Another study found that when frustration and/or confusion transition to negative emotional states, it might be detrimental to learning outcomes (Di Leo et al., 2019).

One of the common factors between frustration and confusion is that even though they both can be intuitively considered detrimental to learning, students might be actively engaging to reduce the duration of these two affective states (Arguel, Lockyer, Kennedy, Lodge, & Pachman, 2019; D'Mello & Graesser, 2012; D'Mello et al., 2014). Boredom, on the other hand, can be linked with disengagement from the learning context/content (D'Mello & Graesser, 2012). Boredom can be the most difficult to avoid once the learners start experiencing it (Arguel et al., 2019; Baker et al., 2010; Csikszentmihalyi, 1997), and therefore it can easily result in undesired learning outcomes (Pekrun, 2006). Furthermore, unlike frustration and confusion, boredom has been found to be negatively correlated in multiple educational settings. For example, in an English as a foreign language class, Kruk found a negative correlation between boredom and learning performance (Kruk, 2016). Other similar examples of learning scenarios where the authors reported a negative correlation between frustration and learning outcomes are language learning (Kruk, 2021), exams (Carlsson, Winder, Eriksson, & Wallerstedt, 2020), and online learning (Heckel & Ringeisen, 2017).

When students collaborate in front of a computer, accomplishing a coding task (co-located collaboration), there is a certain level of social engagement, and a common goal is the creation of a functioning artifact (Papavlasopoulou, Giannakos, & Jaccheri, 2019). An important issue to consider is to keep acceptable levels of participation and strong relationships while students' collaborate (Kwon, Liu, & Johnson, 2014). The associated interactions with these aspects of the group performance can be characterized as social-emotional interactions (Kwon et al., 2014), and these are primarily directed towards the relationship between group members (Isohätälä, Näykki, & Järvelä, 2020). When students collaborate, they must maintain durable relationships and acceptable levels of participation. Interactions associated with these aspects of the group performance can be typified as socio-emotional interactions (Van Diggelen & Overdijk, 2007). These interactions are primarily directed toward the relationship between group members (Dore, 2016). Regarding collaborative learning, positive emotions were correlated with effort and persistence, while negative emotions were correlated with less risk tolerance, lower learning gains, and conflicts (Dore, 2016; Linnenbrink-Garcia, Rogat, & Koskey, 2011). Furthermore, negative socio-emotional interactions such as lack of respect and excessive criticism have significant consequences on the general quality of group learning opportunities (Lescano & Costaguta, 2018) since such groups were reported to undermine commitment (Linnenbrink-Garcia et al., 2011) and criticism (Lajoie et al., 2015).

During the collaboration, confusion occurs when the groups have to reinforce their pre-existing mental models with new information (Clarebout & Elen, 2001; D'Mello & Graesser, 2012). On the other hand, frustration during collaborative learning sessions was found to be eminent during online interaction (Capdeferro & Romero, 2012), and online discussion forums (Chen & Caropreso, 2004). Frustration and confusion were shown to lead to impasses in collaborative learning (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). Lastly, when the problem at hand is far too easy or repetitive, boredom is the emotion that is mainly observed (Panitz, 1999); and the same happens with individual learning (Csikszentmihalyi, 1997). Based on a selective meta-analysis with 21 studies (D'Mello, 2013), in this paper, we decided to focus on these three emotions along with delight and neutral because they were found to be most prominent. Similarly, from the CVT-based set of emotions, fear and disgust might not be particularly related to learning (Arguel et al., 2019).

3. Methodology

3.1. The coding workshop

Our coding activity at the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, is designed based on the constructionist approach and making (Papavlasopoulou et al., 2017). It is a coding workshop in an informal environment at the University's premises and in specially designed rooms. School classes from the region are invited to participate in a one-day out-of-school activity. The workshop's main goal is to introduce CS and programming to children playfully and interactively and does not require any previous coding experience from them. The total duration of the workshop is 4 h and has two sessions. Especially children 13–16 years old are introduced to block-based programming through the Scratch environment. They aim to work collaboratively in dyads or triads; they imagine, create and modify their games by iteratively coding and testing them. Children's teams are instructed by student assistants who show and explain the coding tasks that need to be done to successfully code their games. Although the main instructions for the tasks are the same, children in each team make their own decisions for their games, interact, and discover their knowledge. Therefore, as requested, each of the instructors also provides help to one or two teams during the activities. During the workshops, three researchers are also present to observe, take field notes and ensure its smooth execution. When all teams have completed their games, children shuffle around and play each other's games.

3.2. Sampling and data collection

We collected the data from the coding workshops that happened during Autumn 2017, and children from 8th to 10th grade (age 13–16 years old) participated. The sample comprised 105 participants, 69 boys and 36 girls (mean age: 14.55, SD: 0.650). We collected the videos from ten dyads and ten triads while they were working on coding their games. We had previously collected the necessary consent from the legal guardian for all children, and each child's participation in the study was voluntary. Also, the project is reported to Norwegian Center for Research Data, and all recommendations and regulations for research are followed. The data collection included:

Video recording: To capture children's facial expressions and extract their emotions while coding their game, we used a wide-angle Logitech Webcam. The web camera was placed on the computer. The teams were working and were zoomed at 150% into the children's faces capturing video at 10 FPS. The collected videos were from 50 children (29 females), ten triads, and ten dyads.

Artifacts (the created games): During the coding workshop's process, we collected four versions of the games as artifacts created by each of the teams. Starting from the first version, which was saved 45 min after the start of the workshop, the next game versions were saved every 45 min. This time frame was suggested by the instructors responsible for the coding workshops. They have run them for many years and have gained experience on

Table 1
Group of action units corresponding to each emotions.

Expression	Combination of action units
Boredom	AU4, AU7, AU12
Frustration	AU12, AU43
Confusion	AU1, AU4, AU7, AU12
Delight	AU4, AU7, AU12, AU25,AU26

Table 2
Group of action units corresponding to each emotions.

Emotion	Combination of action units
Happiness	AU6, AU10
Sadness	AU1, AU4, A15
Anger	AU4, AU5, AU7, AU23
Surprise	AU1, AU2, AU5, AU26

how children are experiencing the learning process. Their suggestion derived from defining the best timing for us to monitor children's progress without losing important information about their progress. At the same time not to disturb them or leave too short time between the different versions that would have shown no progress.

Questionnaire: After the end of the workshop, the children completed a paper-based post questionnaire regarding their perceived collaborative learning performance.

3.3. Measurements

We will use two sets of measurements for this paper. First, the duration of the facial expressions that indicate the emotions. Second, the transitions from one to other expressions. Before we explain these two sets of measurements, we will present how we get from the facial videos of the teams to the individual expressions. The action units are basic movements of the facial muscle groups. These action units were originally proposed by Carl-Herman Hjortsjö and were further developed by Paul Ekman, and Wallace Friesen (Eckmann, Kamphorst, & Ruelle, 1987). The action units are the basic building blocks of the Facial Action Coding System. This comprehensive anatomy-based system describes almost all visually perceivable facial activities. Using action units, we can determine which emotions are displayed in every frame of the video where there is a face detected. If we combine certain action units, we can capture certain emotions. The intensity of the action units combined indicates how prominent each detected emotion is. Following are the steps to compute the emotions from the facial video of the collaborating children:

1. Detect the faces in every frame of the video (Fig. 1 left).
2. Align the faces across the frames so that same faces are being tracked and assigned the same ID in every frame by using the method described in Sharma et al. Sharma et al. (2019) (Fig. 1 right).
3. Once we have the faces with correct IDs, use OpenFace (Amos, Ludwiczuk, Satyanarayanan, et al., 2016; Baltrušaitis, Robinson, & Morency, 2016) to compute the Action Units (AUs) (Hager, 2002) for each frame (Fig. 2 left).
4. From the AUs compute the probabilities of the five education specific emotions: frustration, boredom, confusion, delight, and neutral (McDaniel et al., 2007) for every frame of the video.
5. From the same AUs compute the probabilities of the five CVT specific emotions: happiness, sadness, anger, surprise, and neutral (McDaniel et al., 2007) for every frame of the video.

3.3.1. Emotions' durations

Once we have the facial action units from the video for each child in the study, we then computed the proportion of time they displayed each of the five Expressions: confusion, boredom, delight, frustration, and neutral. We used the combination of action units to compute individual expressions (inspired by McDaniel et al. (2007)) shown in the Table 1:

From the same action units, we also computed the control value theoretic emotions: happiness, sadness, anger, fear, surprise, disgust, contempt, and neutral. Out of these emotions, fear, disgust, and contempt account for a total of 3.5% time the children were working on collaborative coding. Therefore, for this paper, we decided to use only five emotions: happiness, sadness, anger, surprise, and neutral. We used the combination of action units to compute individual expressions (inspired by McDaniel et al. (2007)) shown in the Table 2. We used only the subset of all the emotions described in the two respective theoretical frameworks because these emotions, in the recorded data, cover more than 95% of the total interaction time and others contribute to the remaining period.

3.3.2. Transitions among the emotions

The second set of measurements was the transition probabilities from one expression to another. The typical transitions are shown in the bottom panels of Fig. 2. We did not consider self-loops in this paper because we are already using proportion of duration of each emotion as first set of measurements.

3.4. Dependent variables

3.4.1. Objective performance — coding performance

We computed coding performance from them every 45 mins' collected artifacts (Scratch code), monitoring their progress. We used a tool called DrScratch (Moreno-León, Robles, & Román-González, 2015). DrScratch has often been used to analyze Scratch projects because it gives a detailed analysis and, at the same time, supports the assessment of computational thinking (CT) skills, using seven CT components: parallelism, logic, flow control, data representation, abstraction, user interactivity, and synchronization. DrScratch is an automatic, easy, and quick way to analyze Scratch projects offering feedback using a score (Moreno-León et al., 2015) and its results indicate comparable assessment with one of a human expert (Moreno-León, Román-González, Harteveld, & Robles, 2017). Troiano et al. 2019 used DrScratch to examine the progress of each CT component while students were designing their games. Troiano et al. (2019).

Our collected projects (i.e., the four versions of the games created by each team) were uploaded and analyzed by DrScratch online. The results gave us a general score for the project (i.e., max 21), computed by summing up the individual scores the project gets at each of the seven CT components (i.e., from 1 to 3). Fig. 3 shows two examples from the analysis. For the rest of this paper, we will refer to "coding performance measure" as "performance score." We continued our analysis of the four performance scores by using a median cut to split the children's teams into high and low-performing groups for all the phases. The medians for the four phases were 6, 10, 12.5, and 13, respectively. **We labeled a group "high" performing if in at least two out of the four phases the team had higher than the median points for that particular phase. Otherwise, the team was labeled as "low" performing.**

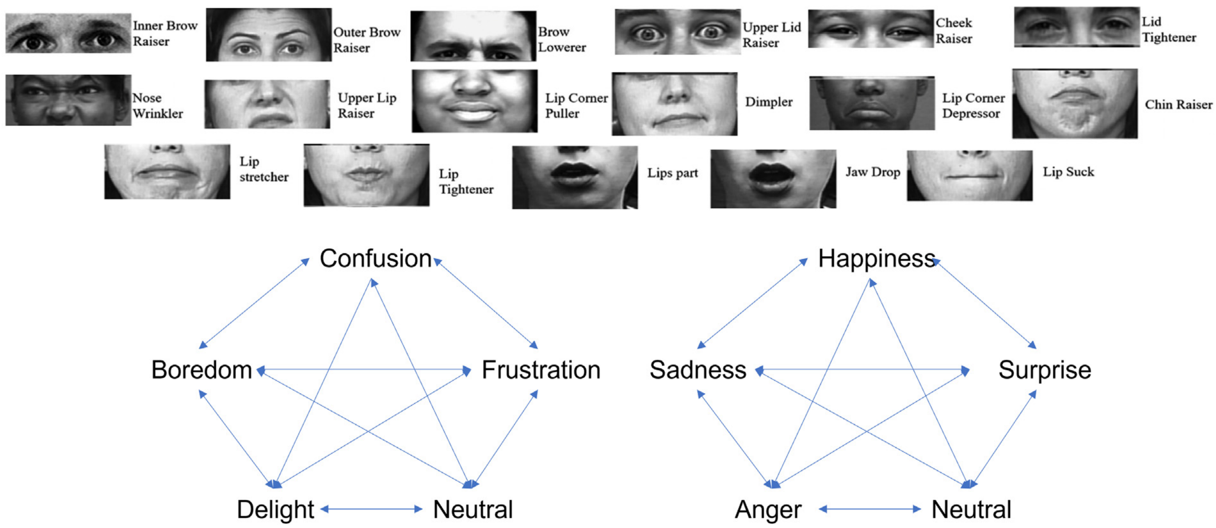

Fig. 1. Top panel: Action units that could be detected. Bottom panel (left): Typical transition diagram for the education specific emotions. Bottom panel (right): Typical transition diagram for the Control Value Theory(CVT) specific emotions.

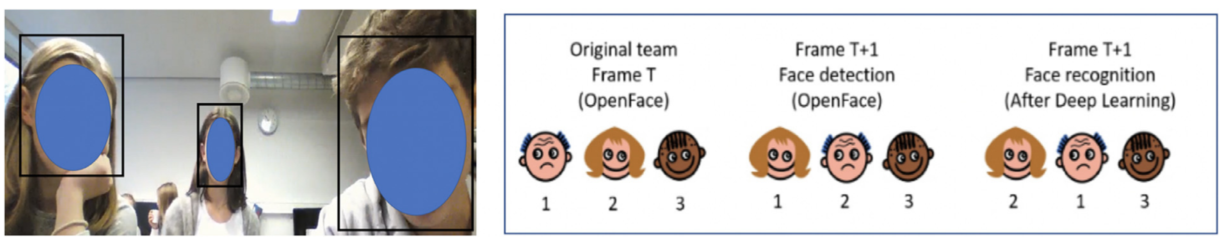

Fig. 2. Left: Example of multiple face detection in one frame. Right: Mitigation scheme for countering the movement of the children during the coding workshop.

Fig. 3. Two examples from the second (left) and the third (right) phases of the coding activity.

3.4.2. Subjective performance — perceived performance

At the end of the activity, the children completed a paper-based survey. The surveys gathered feedback on the children's attitudes regarding the collaborative coding activity. The children were asked to rate their experience with the collaborative coding activity regarding their perceived learning (So & Brush, 2008). In all questions, a seven-point Likert scale. Perceived learning (we refer to this as subjective performance in this paper) is the degree to which children indicate their performance based on the collaborative activity and the outputs they had seen during the coding

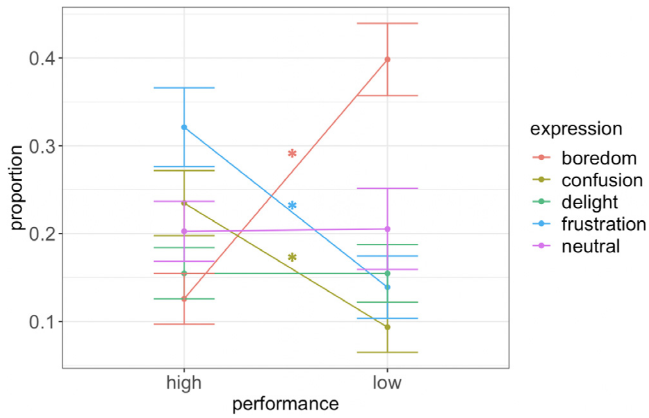

Fig. 4. Comparing the proportional duration of the expressions for the two levels of performance (high/low). The asterisk show the significant differences. The vertical bars are the 95% confidence interval.

activities. This measurement for subjective performance has been used in CCI studies (Papavlasopoulou, 2019; Papavlasopoulou, Sharma, & Giannakos, 2018; Tisza & Markopoulos, 2021). For the groups' subjective performance, we took the mean of all the ratings from the individual members of the group. Once we had every group's perceived performance, we used a median split on the groups' scores to categorize the teams into high and low-performing ones.

3.5. Data analysis

To answer the first research question (**relation between the expressions and children's performance**), we use a t-test with the duration of expressions as the dependent variable and the performance levels (high/low) as the independent variable. Further, to answer the second research question (**how do the expressions change during the coding activity**), we use a t-test with the transition among expressions as the dependent variable and the performance levels (high/low) as the independent variable. For testing the normality, we used the Shapiro–Wilk test (Royston, 1982), and for testing the homoscedasticity, we used the Breusch–Pegan test (Breusch & Pagan, 1979). To account for the multiple comparisons, we applied Bonferroni corrections to get the corrected p-values.

4. Results

4.1. Objective performance and education specific emotions

Fig. 4 shows the comparison of expressions' proportions between individuals from high and low performing teams, and Table 3 shows the mean, standard deviations, and the t-test results. We observe that the individuals in the high performing groups show significantly higher proportions of confusion ($t(41.09) = 5.81$, $p < .00001$) and frustration ($t(41.74) = 6.13$, $p < .00001$) than those from the individuals in the low performing groups. On the other hand, the individuals in the high performing groups show significantly lower proportions of boredom ($t(46.45) = -10.65$, $p < .00001$) than the boredom displayed by the individuals in the low performing groups. Finally, we did not find any difference in the proportions of neutral and delight between the individuals from high and low performing teams (Table 3, Fig. 4).

Fig. 5 and Table 3 show the results for comparing the transitions among the expressions, i.e., confusion, boredom, delight, neutral and frustration, for the two levels of performance (high/low).

Regarding the transitions from the confusion (Fig. 5, top-left), we observe that the individuals from the high performing teams move from confusion to delight ($t(42.86) = 5.79$, $p < .000001$) and neutral ($t(47.82) = 10.73$, $p < .000001$) significantly more than the individuals from the low performing teams. On the other hand, the individuals from the low-performing teams move from confusion to boredom ($t(40.61) = -15.10$, $p < .000001$) significantly more than the individuals from the low-performing teams. There is no difference in moving from confusion to frustration based on the performance levels.

When it comes to the transitions from the frustration, (Fig. 5, top-right), we observe that the individuals from the high performing teams move from frustration to delight ($t(47.81) = 6.43$, $p < .000001$) and neutral ($t(47.85) = 8.97$, $p < .000001$) significantly more than the individuals from the low performing teams. On the other hand, the individuals from the low-performing teams move from frustration to boredom ($t(47.84) = -13.21$, $p < .000001$) significantly more than the individuals from the low-performing teams. There is no difference in moving from frustration to confusion based on the performance levels.

Considering the transitions from the boredom, (Fig. 5, middle), we observe that the individuals from the high performing teams move from boredom to neutral ($t(42.41) = 16.46$, $p < .000001$) significantly more than the individuals from the low performing teams. On the other hand, the individuals from the low-performing teams move from boredom to frustration ($t(46.99) = -26.26$, $p < .000001$) significantly more than the individuals from the low-performing teams. There is no difference in moving from boredom to confusion and boredom to delight based on the performance levels.

Concerning the transitions from the delight, (Fig. 5, bottom-left), we observe that the individuals from the high performing teams move from delight to confusion ($t(47.78) = 4.23$, $p < .0001$) and frustration ($t(47.92) = 3.40$, $p < .001$) significantly more than the individuals from the low performing teams. On the other hand, the individuals from the low-performing teams move from delight to boredom ($t(46.62) = -11.78$, $p < .000001$) significantly more than the individuals from the low-performing teams. There is no difference in moving from delight to neutral based on the performance levels.

4.2. Objective performance and CVT specific emotions

Next, we analyzed the differences between the proportions of CVT-specific emotions and the transitions among them across the two performance levels. From Table 4 and Fig. 6, we observe that the high performing teams show significantly lower levels of sadness ($t(36.43) = -3.48$, $p = .001$) and anger ($t(47.92) = -10.04$, $p = .00001$) during the coding activities than the low performing teams. There were no other differences based on the proportions of the CVT-based emotions between the high and low-performing teams.

Considering the transitions from happiness, we observe that the high performing teams transition to sadness ($t(46.79) = -3.80$, $p = .0004$) and anger ($t(32.13) = -8.26$, $p = .00001$) with a significantly lower probability than the low performing teams. Moreover, the high performing teams transition from sadness to anger ($t(32.55) = -4.64$, $p = .00001$) and anger to sadness ($t(39.48) = -3.09$, $p = .003$) with a significantly lower probability than the low performing teams. For all the other transitions between CVT-based emotions, we did not find any significant differences between the high performing teams and the low performing teams, when it comes to the objective performance (Fig. 7).

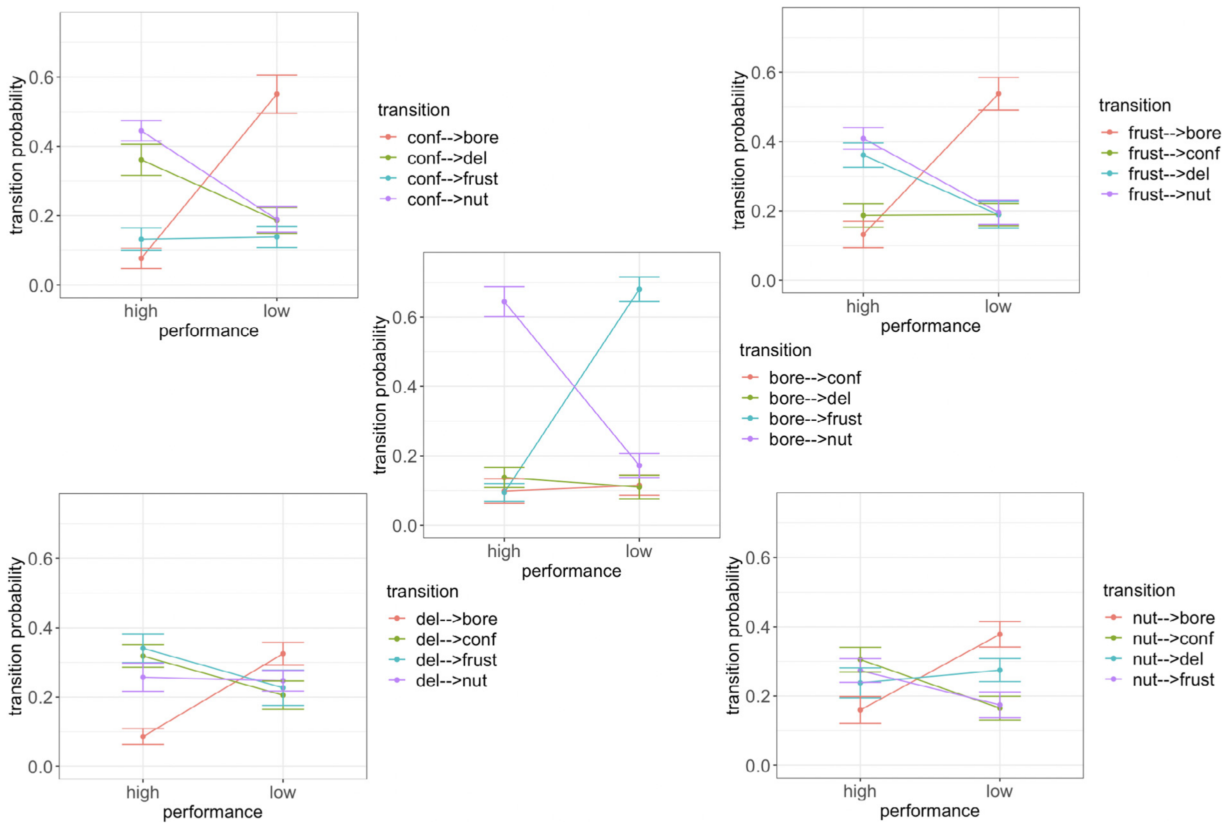

Fig. 5. Comparing the transitions among the expressions for the two levels of performance (high/low). **Top-left:** transitions from confusion; **Top-right:** transitions from frustration; **Middle:** transitions from boredom; **Bottom-left:** transitions from delight; **Bottom-right:** transitions from neutral; **conf = confusion; frust = frustration; nut = neutral; bore = boredom; del = delight**. The vertical bars are the 95% confidence interval.

Table 3
Comparing the proportional duration of the expressions and transitions among the expressions for the two levels of objective performance (high/low). All the mean, SD, and t-values are rounded to two significant digits. For consistency of effect sizes, all the effect sizes are calculated with a degree of freedom = 48. This is the ceiling of the maximum degree of freedom in this contribution. **con = confusion; fru = frustration; nut = neutral; bor = boredom; del = delight; SD = standard deviation**.

	Mean (SD)	Mean (SD)	t-value	p-value	Effect size		Mean (SD)	Mean (SD)	t-value	p-value	Effect size
	High	**Low**				**Fru-Nut**	0.40 (0.03)	0.19 (0.03)	8.97	0.000001	1.29
Boredom	0.12 (0.02)	0.39 (0.04)	−10.65	0.00001	1.53	**Con-Bor**	0.07 (0.02)	0.51 (0.05)	−15.10	0.000001	2.16
Frustration	0.32 (0.04)	0.13 (0.03)	6.13	0.00001	0.88	**Con-Fru**	0.13 (0.03)	0.13 (0.03)	−0.29	0.76	0.04
Confusion	0.23 (0.03)	0.09 (0.02)	5.81	0.00001	0.84	**Con-Del**	0.36 (0.04)	0.18 (0.03)	5.79	0.000001	0.83
Delight	0.15 (0.02)	0.15 (0.03)	0.003	0.99	0.00	**Con-Nut**	0.44 (0.02)	0.18 (0.03)	10.73	0.000001	1.53
Neutral	0.20 (0.03)	0.20 (0.04)	−0.008	0.92	0.01	**Del-Bor**	0.08 (0.02)	0.24 (0.03)	−11.78	0.000001	1.70
Bor-Fru	0.09 (0.02)	0.68 (0.03)	−26.26	0.000001	3.77	**Del-Fru**	0.34 (0.04)	0.22 (0.05)	3.40	0.001	0.49
Bor-Con	0.09 (0.03)	0.11 (0.02)	−0.69	0.48	0.10	**Del-Con**	0.31 (0.03)	0.20 (0.04)	4.23	0.0001	0.61
Bor-Del	0.13 (0.02)	0.11 (0.03)	1.21	0.23	0.17	**Del-Nut**	0.25 (0.04)	0.24 (0.02)	0.38	0.69	0.05
Bor-Nut	0.64 (0.04)	0.17 (0.03)	16.16	0.000001	2.38	**Nut-Bor**	0.15 (0.03)	0.37 (0.03)	−7.91	0.000001	1.14
Fru-Bor	0.13 (0.03)	0.53 (0.04)	−13.21	0.000001	1.95	**Nut-Fru**	0.27 (0.03)	0.17 (0.03)	3.83	0.0003	0.55
Fru-Con	0.18 (0.03)	0.18 (0.03)	−0.10	0.91	0.01	**Nut-Con**	0.30 (0.03)	0.16 (0.03)	5.50	0.000001	0.79
Fru-Del	0.36 (0.03)	0.18 (0.03)	6.43	0.000001	0.92	**Nut-Del**	0.23 (0.03)	0.27 (0.03)	−1.30	0.20	0.18

Table 4
Comparing the proportional duration of the CVT-based emotions and transitions among the expressions for the two levels of objective performance (high/low). All the mean, SD and t-values are rounded to two significant digits. For consistency of effect sizes, all the effect sizes are calculated with degree of freedom = 48. This is the ceiling of the maximum degree of freedom in this contribution. **hap = happiness; sad = sadness; nut = neutral; ang = anger; sup = surprise; SD = standard deviation**.

	Mean (SD)	Mean (SD)	t-value	p-value	Effect Size		Mean (SD)	Mean (SD)	t-value	p-value	Effect size
	High	**Low**				**Sad-Nut**	0.40 (0.14)	0.34 (0.14)	1.68	0.09	0.23
Happiness	0.23 (0.1)	0.19 (0.1)	−0.14	0.67	0.02	**Ang-Hap**	0.21 (0.12)	0.25 (0.12)	−1.57	0.12	0.22
Sadness	0.16 (0.05)	0.23 (0.05)	−3.48	0.001	0.44	**Ang-Sad**	0.32 (0.1)	0.21 (0.1)	−3.09	0.003	0.40
Anger	0.19 (0.01)	0.22 (0.01)	−10.04	0.00001	0.82	**Ang-Sup**	0.25 (0.13)	0.26 (0.13)	−0.49	0.61	0.07
Surprise	0.15 (0.03)	0.14 (0.04)	1.67	0.11	0.23	**Ang-Nut**	0.22 (0.14)	0.27 (0.14)	0.28	0.77	0.04
Neutral	0.23 (0.1)	0.17 (0.1)	1.28	0.20	0.18	**Sup-Hap**	0.35 (0.1)	0.33 (0.1)	0.64	0.52	0.09
Hap-Sad	0.30 (0.03)	0.35 (0.04)	−3.80	0.0004	0.48	**Sup-Sad**	0.21 (0.1)	0.23 (0.1)	−0.58	0.62	0.08
Hap-Ang	0.20 (0.02)	0.25 (0.02)	−8.26	0.00001	0.77	**Sup-Ang**	0.25 (0.1)	0.23 (0.1)	1.15	0.21	0.16
Hap-Sup	0.30 (0.1)	0.27 (0.1)	0.82	0.41	0.11	**Sup-Nut**	0.19 (0.1)	0.21 (0.1)	−0.78	0.43	0.11
Hap-Nut	0.20 (0.12)	0.13 (0.12)	1.20	0.23	0.17	**Nut-Hap**	0.30 (0.13)	0.27 (0.13)	1.45	0.15	0.20
Sad-Hap	0.18 (0.12)	0.16 (0.12)	1.30	0.19	0.18	**Nut-Sad**	0.30 (0.1)	0.28 (0.1)	−0.37	0.71	0.05
Sad-Ang	0.22 (0.1)	0.35 (0.1)	−4.64	0.00001	0.55	**Nut-Ang**	0.15 (0.12)	0.20 (0.12)	1.15	0.26	0.16
Sad-Sup	0.18 (0.1)	0.15 (0.1)	1.13	0.26	0.16	**Nut-Sup**	0.25 (0.1)	0.25 (0.1)	−0.78	0.43	0.11

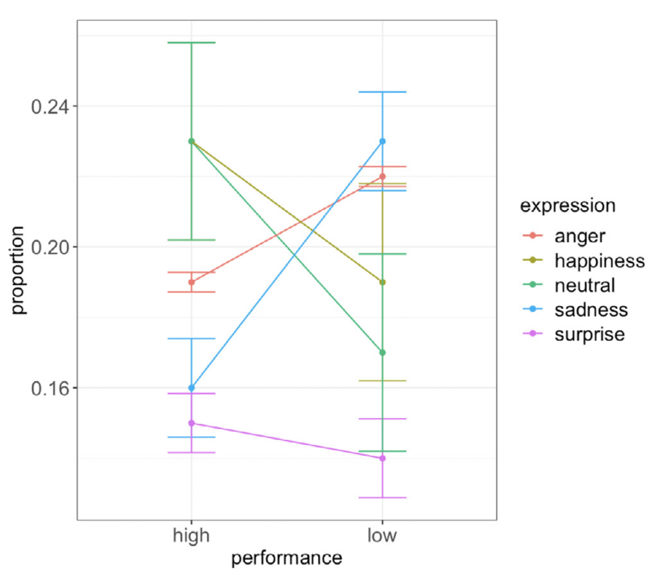

Fig. 6. Comparing the proportional duration of the CVT-based expressions for the two levels of objective performance (high/low). The vertical bars are the 95% confidence interval.

4.3. Subjective performance and education specific emotions

When we analyzed the relationships between the education-specific emotions and the subjective performance levels (high/low), we did not. observe any significant differences between the proportions of the individual emotions (i.e., boredom, frustration, confusion, delight, and neutral) across the high and low-performing teams. Furthermore, we also did not find any significant differences between the transitions for high performing teams and the transitions for the low performing teams when the subjective performance was considered (Table 5, Figs. 8 and 9).

4.4. Subjective performance and CVT specific emotions

Finally, we analyzed the relationships between the CVT based emotions and the subjective performance levels (high/low). We observe, from Table 6 and Fig. 11, that high performing teams display significantly higher levels of happiness ($t(42.39) = 4.49$, $p = .00001$) during the coding activities than the low performing teams. Furthermore, the high performing teams also show significantly lower levels of sadness ($t(41.48) = -3.72$, $p = .0006$) and anger ($t(40.36) = -18.07$, $p = .00001$) than the low performing teams (see Fig. 10). Analyzing the transitions between the CVT-based emotions and across the two levels of subjective performance (Fig. 11), we observe that there are significant differences between the high and low performing teams based on the transitions between sadness, happiness, and anger. However, there are no other significant differences between the high and low-performing teams when it comes to any other transition.

Concerning the transitions among happiness, sadness and anger, we observe that the high performing teams have significantly higher transitions from sadness to happiness ($t(38.71) = 6.50$, $p = .00001$) and from anger to happiness ($t(47.23) = -4.94$, $p = .00001$) than the low performing teams. On the other hand, the high performing teams have significantly lower transitions from sadness to anger ($t(47.11) = -2.73$, $p = .01$), anger to sadness ($t(45.32) = -3.09$, $p = .003$), happiness to sadness ($t(44.52) = -2.38$, $p = .02$), and happiness to anger ($t(39.05) = -2.58$, $p = .01$) than the low performing teams.

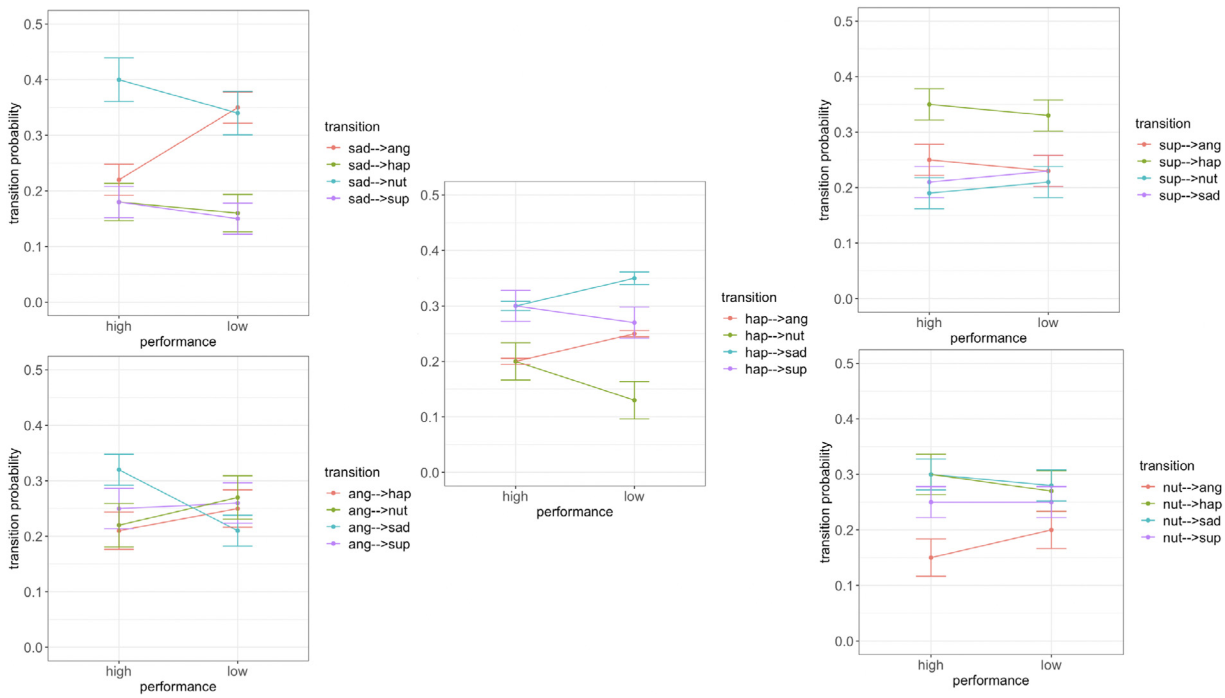

Fig. 7. Comparing the transitions among the CVT-based expressions for the two levels of objective performance (high/low). **Top-left:** transitions from sadness; **Top-right:** transitions from surprise; **Middle:** transitions from happiness; **Bottom-left:** transitions from anger; **Bottom-right:** transitions from neutral; **conf = confusion; frust = frustration; nut = neutral; bore = boredom; del = delight**. The vertical bars are the 95% confidence interval.

Table 5
Comparing the proportional duration of the expressions and transitions among the expressions for the two levels of subjective performance (high/low). All the mean, SD and t-values are rounded to two significant digits. For consistency of effect sizes, all the effect sizes are calculated with degree of freedom = 48. This is the ceiling of the maximum degree of freedom in this contribution. **con = confusion; fru = frustration; nut = neutral; bor = boredom; del = delight; SD = standard deviation**.

	Mean (SD)	Mean (SD)	t-value	p-value	Effect size		Mean (SD)	Mean (SD)	t-value	p-value	Effect size
	High	**Low**				**Fru-Nut**	0.39 (0.11)	0.41 (0.11)	0.08	0.93	0.01
Boredom	0.18 (0.11)	0.19 (0.11)	−0.81	0.42	0.11	**Con-Bor**	0.20 (0.13)	0.21 (0.13)	−0.46	0.64	0.06
Frustration	0.11 (0.12)	0.12 (0.12)	−0.15	0.87	0.02	**Con-Fru**	0.15 (0.11)	0.16 (0.11)	1.39	0.17	0.19
Confusion	0.16 (0.12)	0.15 (0.11)	−0.28	0.77	0.04	**Con-Del**	0.37 (0.12)	0.33 (0.12)	1.18	0.24	0.04
Delight	0.21 (0.13)	0.22 (0.13)	0.37	0.70	0.05	**Con-Nut**	0.25 (0.12)	0.26 (0.12)	−1.39	0.16	0.19
Neutral	0.19 (0.12)	0.19 (0.12)	0.07	0.94	0.01	**Del-Bor**	0.25 (0.1)	0.26 (0.1)	0.90	0.37	0.12
Bor-Fru	0.12 (0.1)	0.11 (0.1)	1.38	0.17	0.19	**Del-Fru**	0.25 (0.1)	0.24 (0.1)	−0.93	0.35	0.13
Bor-Con	0.21 (0.1)	0.2 (0.1)	−1.20	0.23	0.17	**Del-Con**	0.34 (0.11)	0.35 (0.11)	−1.92	0.06	0.26
Bor-Del	0.12 (0.12)	0.14 (0.12)	−1.16	0.24	0.16	**Del-Nut**	0.12 (0.11)	0.10 (0.11)	1.30	0.19	0.18
Bor-Nut	0.55 (0.1)	0.53 (0.1)	0.30	0.76	0.04	**Nut-Bor**	0.10 (0.13)	0.08 (0.13)	−0.23	0.81	0.03
Fru-Bor	0.1 (0.11)	0.1 (0.11)	0.41	0.65	0.05	**Nut-Fru**	0.32 (0.1)	0.30 (0.1)	0.10	0.92	0.01
Fru-Con	0.17 (0.13)	0.15 (0.13)	0.50	0.62	0.07	**Nut-Con**	0.31 (0.12)	0.33 (0.12)	−0.07	0.93	0.01
Fru-Del	0.25 (0.12)	0.28 (0.12)	−1.58	0.12	0.22	**Nut-Del**	0.21 (0.1)	0.22 (0.1)	−1.91	0.06	0.26

Table 6
Comparing the proportional duration of the CVT-based emotions and transitions among the expressions for the two levels of subjective performance (high/low). All the mean, SD and t-values are rounded to two significant digits. For consistency of effect sizes, all the effect sizes are calculated with degree of freedom = 48. This is the ceiling of the maximum degree of freedom in this contribution. **hap = happiness; sad = sadness; nut = neutral; ang = anger; sup = surprise; SD = standard deviation**.

	Mean (SD)	Mean (SD)	t-value	p-value	Effect Size		Mean (SD)	Mean (SD)	t-value	p-value	Effect size
	High	**Low**				**Sad-Nut**	0.21 (0.13)	0.22 (0.13)	−0.34	0.73	0.04
Happiness	0.25 (0.1)	0.17 (0.1)	4.49	0.00001	0.54	**Ang-Hap**	0.31 (0.06)	0.25 (0.06)	4.94	0.00001	0.58
Sadness	0.16 (0.05)	0.22 (0.05)	−3.72	0.0006	0.47	**Ang-Sad**	0.21 (0.1)	0.32 (0.1)	−3.09	0.003	0.40
Anger	0.17 (0.01)	0.22 (0.01)	−18.07	0.00001	0.93	**Ang-Sup**	0.21 (0.13)	0.23 (0.13)	−0.63	0.52	0.09
Surprise	0.15 (0.03)	0.15 (0.04)	0.20	0.83	0.02	**Ang-Nut**	0.17 (0.14)	0.20 (0.14)	0.82	0.41	0.11
Neutral	0.23 (0.1)	0.21 (0.1)	0.87	0.39	0.12	**Sup-Hap**	0.25 (0.1)	0.23 (0.1)	0.85	0.39	0.12
Hap-Sad	0.28 (0.1)	0.32 (0.1)	−2.38	0.02	0.32	**Sup-Sad**	0.31 (0.1)	0.33 (0.1)	−1.39	0.17	0.19
Hap-Ang	0.18 (0.12)	0.25 (0.12)	−2.58	0.01	0.34	**Sup-Ang**	0.35 (0.1)	0.33 (0.1)	1.63	0.11	0.22
Hap-Sup	0.19 (0.11)	0.21 (0.11)	−0.19	0.884	0.02	**Sup-Nut**	0.09 (0.1)	0.11 (0.1)	−0.10	0.91	0.01
Hap-Nut	0.19 (0.1)	0.21 (0.1)	0.10	0.92	0.01	**Nut-Hap**	0.33 (0.1)	0.34 (0.1)	−1.85	0.15	0.25
Sad-Hap	0.36 (0.04)	0.20 (0.04)	6.50	0.00001	0.68	**Nut-Sad**	0.33 (0.1)	0.32 (0.1)	−0.58	0.40	0.08
Sad-Ang	0.25 (0.1)	0.33 (0.1)	−2.73	0.01	0.36	**Nut-Ang**	0.15 (0.1)	0.17 (0.1)	−0.40	0.68	0.05
Sad-Sup	0.16 (0.12)	0.15 (0.12)	0.22	0.82	0.03	**Nut-Sup**	0.18 (0.1)	0.17 (0.1)	0.62	0.53	0.08

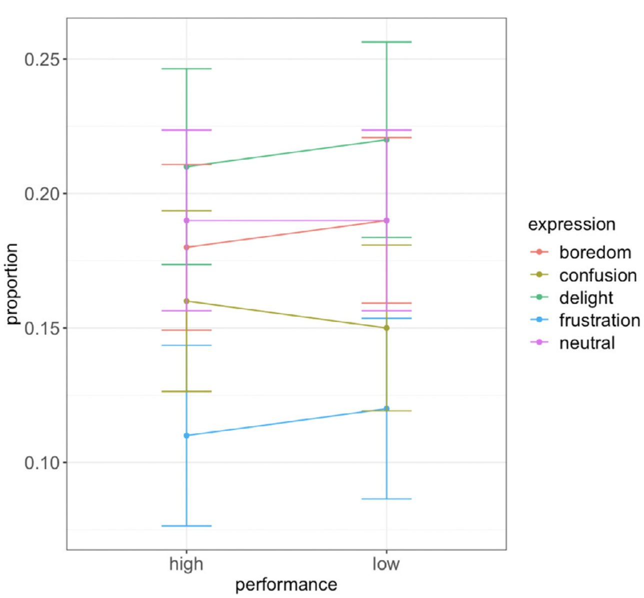

Fig. 8. Comparing the proportional duration of the education specific expressions for the two levels of subjective performance (high/low). The vertical bars are the 95% confidence interval.

5. Discussion

We observe from the analysis that there are clear differences between the individuals from the objectively high-performing teams and the objectively low-performing teams on account of the durations and transitions among the education-specific emotions (Research Question 1). There were certain (and fewer) differences between the two objective performance levels based on the durations and transitions among the Control Value Theory (CVT) specific emotions (Research Question 2). We also see a clear difference between the subjectively high-performing and low-performing teams. We observe such differences while analyzing the durations and transitions among the Control Value Theory (CVT) specific emotions (Research Question 4). In contrast, we did not find any relationships between the subjective performance levels and the education-specific emotions (Research Question 3). In this paper, we chose to utilize the education theoretic emotions (i.e., frustration, boredom, confusion, delight, neutral) and the control value theoretic emotions (i.e., happiness, sadness, anger, surprise, neutral). The main reason for this decision was that the education theoretic emotions are increasingly being used more and more in the past few years in related fields of Learning Analytics (LAK) (Kostyuk, Almeda, & Baker, 2018), User Modeling (UMUAI) (Richey et al., 2019). Another reason was that there are also a few direct relations between the control value theoretic emotions and the task-based performance (Sharma et al., 2019), or academic performance in general (Bless, 2000; Fredrickson, 1998). In this section, we will provide plausible explanations for the results presented in the results section. Moreover, we will also provide implications from practical and research points of view.

5.1. Interpretation of the results

The first research question caters to the difference between the duration of and the transitions among the education-specific emotions (i.e., frustration, boredom, confusion, delight, neutral) shown by the individual team members from the high and low-performing teams from an objective measurement (task-based

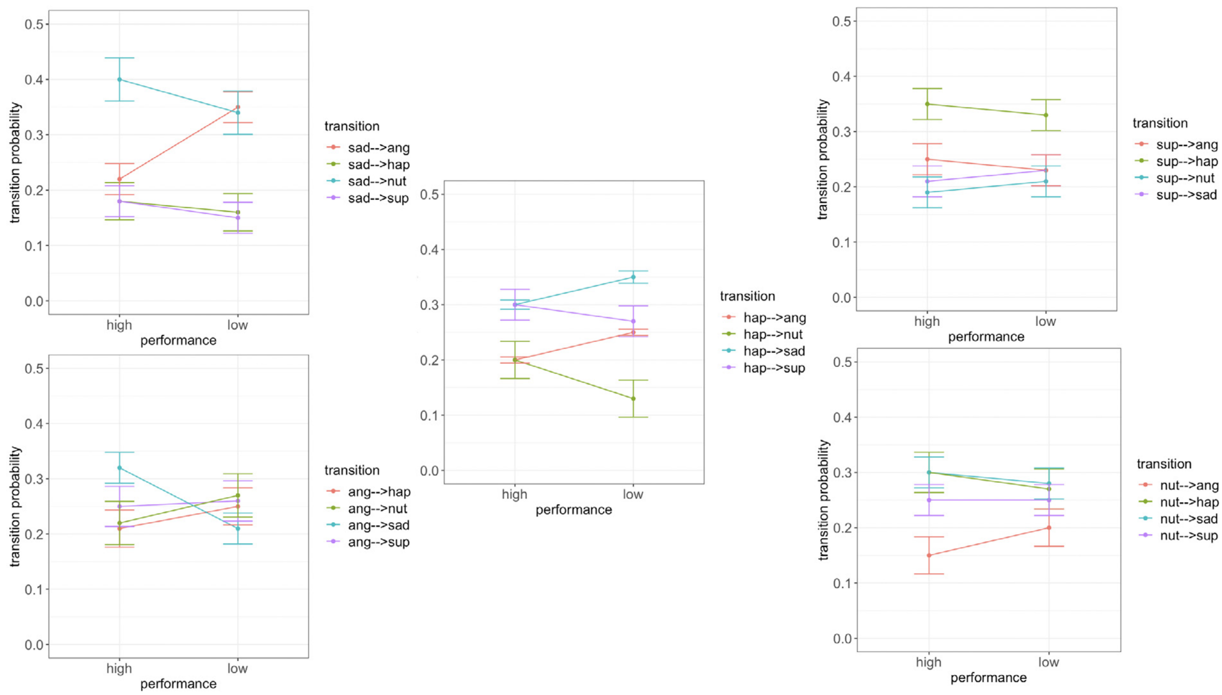

Fig. 9. Comparing the transitions among the education specific expressions for the two levels of subjective performance (high/low). **Top-left:** transitions from frustration; **Top-right:** transitions from delight; **Middle:** transitions from boredom; **Bottom-left:** transitions from confusion; **Bottom-right:** transitions from neutral; **conf = confusion; frust = frustration; nut = neutral; bore = boredom; del = delight**. The vertical bars are the 95% confidence interval.

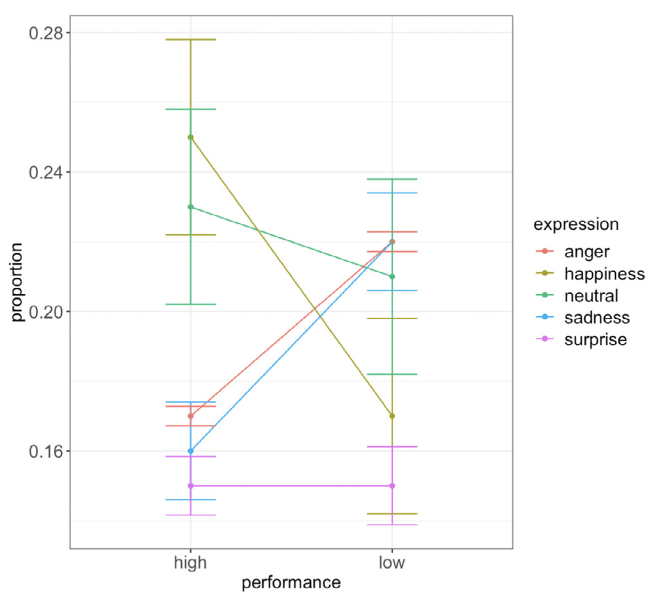

Fig. 10. Comparing the proportional duration of the CVT based expressions for the two levels of subjective performance (high/low). The vertical bars are the 95% confidence interval.

performance). The results show that the individuals from the high-performing teams show more confusion and frustration while individuals from the low-performing teams show more boredom. The recent results from the individual learning scenarios show that students' boredom could be detrimental to their academic and task-based performance (Baker et al., 2010; Dowd et al., 2015). On the other hand, confusion and frustration could be beneficial for the students' learning outcomes (Makewa et al., 2014; Richey et al., 2019). While collaborating on the given coding problem, the students might enter a behavioral loop in which their previous mental models are challenged by the task and try to understand the problem. This behavioral loop might increase their confusion when the code does not work as per their hypothesis (D'Mello & Graesser, 2010). Similarly, they might try to understand what caused the problem, which can increase their frustration (Dowd et al., 2015). However, in certain cases, the students can also disengage from the problem, raising the levels of boredom to cited2012dynamics. From our results, it appears that individuals from the high-performing teams might get involved with problems and the reasons for them and hence show more confusion and frustration than those from the low-performing teams. Whereas the low-performing teams do not engage in active problem solving and therefore show more boredom than those from the high-performing teams.

Understanding the basic differences in durations of these emotions presents one side of the observations from the study where we are only comparing the emotions' durations across the different levels of performance. However, this does not encompass the transitions among the different emotions. Parts of the research questions look at the basic temporality of the emotions from a Markovian point of view. The first research question also addresses the differences between the individuals from the high and low-performing teams based on the transition among the education-specific emotions. The results show that the high-performing teams move from confusion and frustration to delight and neutral and vice versa, while the low-performing teams move from every emotion to boredom. When combined with the first research question results, these results provide interesting insight into the process. On the one hand, for the high-performing teams, we observe that individuals from these teams move from frustration and confusion to delight and neutral more often than the individuals from the low-performing teams. We also know these individuals have shown more confusion and frustration (from analyzing the durations). This shows that when the high-performing teams try to understand the problem (confusion) and/or find the cause of the problem (frustration), they move to delight and neutral more often than the low-performing teams.

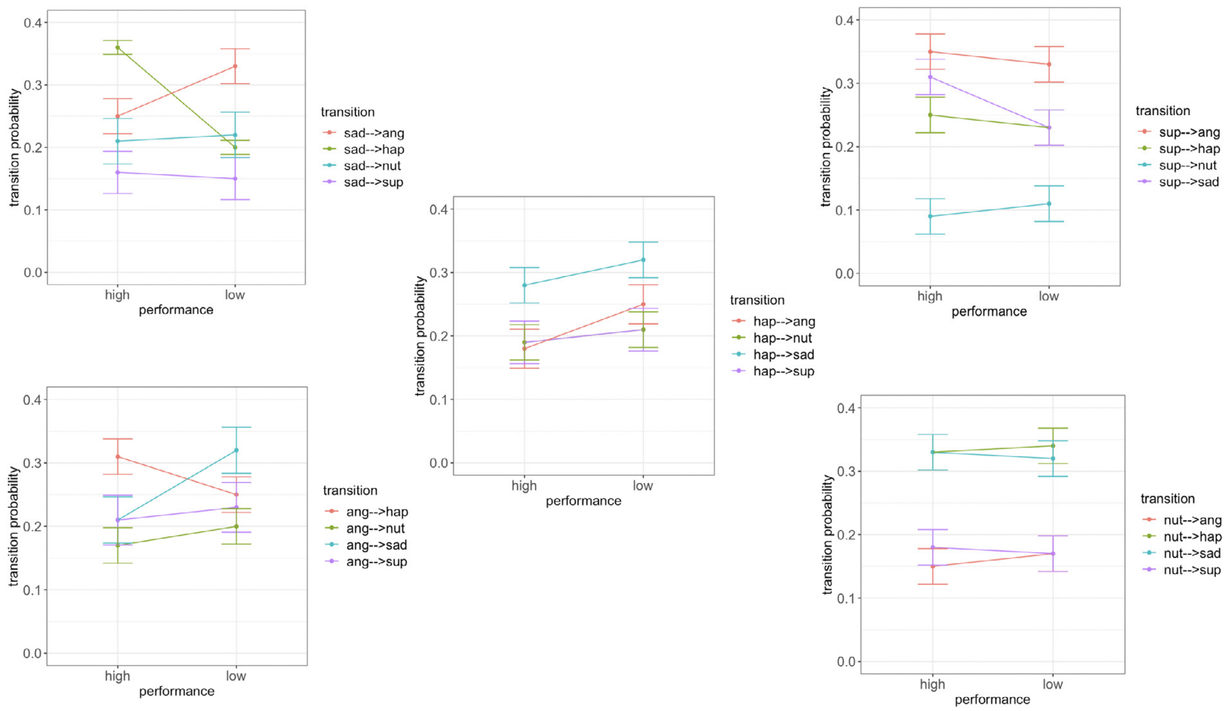

Fig. 11. Comparing the transitions among the CVT based expressions for the two levels of subjective performance (high/low). **Top-left:** transitions from sadness; **Top-right:** transitions from surprise; **Middle:** transitions from happiness; **Bottom-left:** transitions from anger; **Bottom-right:** transitions from neutral; **conf = confusion; frust = frustration; nut = neutral; bore = boredom; del = delight**. The vertical bars are the 95% confidence interval.

The emotions with non-negative connotations are often the results of solving the problem (delight) or having understood the cause of the problem (neutral) (D'Mello & Graesser, 2012). The high-performing teams could be in a similar situation, as reflected in their scores. On the other hand, for the low-performing teams, we observe that the individuals from these teams move to boredom from any other emotion more often than those from the high-performing teams. This shows that the students in the low-performing teams often disengage from the short-term problem-solving processes (Baker et al., 2010; D'Mello & Graesser, 2012), which might lead to low performance.

The second research question seeks to formulate the relationships between the control value theory-based (CVT) emotions and objective performance. There are two key take-away messages for the second research question. First, the emotions with negative valance are predominant in the low-performing teams. Second, there are no apparent differences between high and low objective performance levels based on the emotions with positive valance. There could be two plausible reasons for the fact that while coding, the low-performing teams are feeling more anger and sadness and switching more between these two emotions than the high-performing teams. First, the children in the low-performing teams may engage in coding activities that do not yield their hypothesized results. It has been shown in various domains such as driving (Yip, Wishart, & Barrett, 2020), sports (Schermuly, 2014), games (Barnett, Coulson, & Foreman, 2010), and workplace (Fong & Kleiner, 2004) that certain unwanted or unintended results can lead to emotions like anger and sadness. The second plausible reason could be the introduction of bugs in the code. The children from the teams with low-performance levels might have introduced bugs in the code during the activity and therefore could not obtain the desired results. This second reason is also supported by studies concerning emotions in coding and software engineering (Bosch, D'Mello, & Mills, 2013; Gachechiladze, Lanubile, Novielli, & Serebrenik, 2017; Graziotin, Wang, & Abrahamsson, 2014). These results indicate that positive emotions do not have any relation, but those negative emotions negatively affect the coding performance. Although our results do not suggest causal relations between anger/sadness and objective performance, when combined with certain longitudinal studies from childhood and early-childhood phases, show clear detrimental effects of anger on children's academic performance (Judge & Jahns, 2007; Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017; Zhou, Main, & Wang, 2010). On the other hand, it has also been shown that sadness can have a detrimental effect on academic performance (Flook, Repetti, & Ullman, 2005; Hanson, Austin, & Lee-Bayha, 2004; Kwon, Hanrahan, & Kupzyk, 2017).

The third research question concerns the relationship between education-specific emotions (i.e., frustration, boredom, satisfaction, confusion, and neutral) and subjective performance levels. The results do not indicate any significant differences between the high and low levels of objective performance when it comes to the proportions of emotions or the transitions among them. This indicates that these two constructs might not be related to each other. However, further investigation is necessary for a generalizable claim. Another example is where a set of emotions/expressions motivated by a given theoretical framework do not have a clear relationship with performance measurements. As aforementioned in the related work section, several studies are pointing toward the lack of clear evidence that academic performance and CVT-based emotions are related. This could also be the case with education-specific emotions and subjective performance (perceived performance).

Finally, the fourth research question investigates the relationship between CVT-based emotions (i.e., happiness, sadness, anger, surprise, and neutral) and subjective performance levels. In the case of subjective performance levels, we observe that the positive valance emotion (happiness) has a higher proportion for the teams with high subjective performance and the negative valance emotions (sadness and anger) have higher proportions for the teams with low subjective performance. Moreover, the high-performing teams transition more from negative to positive

Table 7
Mapping between the design considerations and results in this contribution.

Design consideration	Related result
Help children with seeking requests and trigger help from their peers or the instructors	To mitigate boredom and eventual disengagement because boredom is negatively correlated with objective performance
Managing learner's feelings of helplessness	To mitigate prolonged frustration because the transition from frustration to boredom is negatively correlated with objective performance
A more efficient and emotionally-aware team-formation	Transitions confusion to delight and frustration to delight are positively correlated with objective performance
Instructors' actions should respond to the learner's needs accordingly helping them to confront emotional struggles and difficulties	The CVT-based emotions are closely linked with the subjective performance
Design affect-sensitive learning environments	There are clear relations between the emotions and performance
Provide content-based help to the students when the groups are showing confusion and/or frustration	Both are correlated with the objective performance
Provide affective/motivational support to the teams who are displaying more boredom	Boredom is negatively correlated with objective performance
Encouraging children into more playful talk	Sadness and anger are negatively correlated with both objective and subjective performances

valance emotions, and for the low-performing teams, this is the other way around. These results are intuitive and straightforward because the happier the children felt during the coding activity, the higher they rated their perceived performance and vice versa for sadness and anger. Furthermore, the children from the high-performing teams keep returning to their happy state after feeling sad and/or angry more often than those from the low-performing teams. This indicates difficult moments during the collaborative coding sessions, either due to the task at hand or the collaborative activities (e.g., argumentation, negotiation, explanation to peers). The high-performing teams seem to have overcome these difficult moments and transitioned to a state of happiness more often than the low-performing teams. This could be another reason happiness and transitions to happiness were associated with high performance.

5.2. Implications for research and practice

This study is the first step to better understanding children's affective states during coding activities working in teams. To the best of our knowledge, this is the first study that explores the two sets of emotions/expressions/affective states (education-specific and control value theory-based) along with two different performance metrics (objective — task-based performance and subjective — perceived performance). Our results show three clear categories of relationships. First, objective performance is linked with education-specific emotions. Second, the objective performance is linked with the negative valance of the control value theory-based emotions. Third, subjective performance is linked with the control value theory-based emotions. In this section, we will highlight the implications of these categories for both the researchers and practitioners (Table 7 presents the summary of design considerations and respective results).

Our approach will help instructors understand how children face the learning process and gain insights on how to respond to them. For example, help them with seeking requests and trigger help from their peers or the instructors to scaffold their behavior (Israel et al., 2016). Also, this research will help us give more "at the moment" reactions to the interactions that naturally happen during k-12 CS/CT activities (Israel et al., 2016). Benefits vary depending on the specific task and how the group is formed (Barron, 2000). Collaboration is critical for shared engagement in problem-solving and managing learners' feelings of helplessness (Israel et al., 2016). Also, individual characteristics and group dynamics are equally important (Stahl et al., 2014). When children debug, a problem may experience difficulties and need to negotiate their process. For example, in our case, we found that low-performing teams experience more boredom due to poor communication between team members. One child may get control in coding without spending time or effort to involve the other team members in the process, resulting in disengagement and boredom. Such behavior is shown in pair programming in the previous research both with adults (Chang & Tsai, 2018) as well as children (Denner et al., 2014). However, it can be that in high-performing teams, confusion and frustration lead to delight because the team members had different levels of experience before this coding workshop. In a study with Alice programming environment, the higher knowledge gains were for the students with low prior experience in computer use, showing that in pair programming, students who work with someone with more experience can learn (Denner et al., 2014). Pair programming benefits computational thinking, and coding knowledge acquisition, especially for the less experienced students (Denner et al., 2014). Moreover, Rodrigo and Baker 2009, showed that feelings of confusion and boredom were associated with lower achievement in a CS course (Rodrigo et al., 2009). This indicates a more efficient and emotionally aware team formation, as also suggested in a review by Reis and colleagues (Reis et al., 2018).

Although there is intuition on how to help children engage in an effective learning experience in coding tasks, it is useful to have studies that can show how to subjectively extract children's emotions. This can benefit future real-time systems that can support instructors in action by showing them the emotional flow the learners have. Sridhar et al. 2018 stated that there is a need to understand affective states concerning cognitive load. For example, a learner who is curious but remains engaged is different than someone overloaded and anxious due to inability to continue with the tasks (Sridhar et al., 2018). Therefore instructors' actions should respond to the learner's needs accordingly, helping them to confront emotional struggles and difficulties during the learning process. This might also influence the perceived control and value appraisal, as shown by Pekrun (Pekrun et al., 2007). A better understanding of children's affective states during their interaction with coding and working as a team will help us design affect-sensitive learning environments. Those can be systems that may include affective responses into their cycles and help students shift into emotions that will help them facilitate the learning process and have the desired outcomes (Baker et al., 2010).

Instructors and educators can benefit from the knowledge about the relation between the emotional and/or affective processes during collaborative learning settings and the collaborative learning outcome/quality. For example, the instructors and educators can provide content-based help to the students when the groups are showing confusion and/or frustration. In these two cases, the students are either struggling from a mismatch between their knowledge model and the actual content (confusion, D'Mello et al. (2014)) or they are struggling with the content itself because the content is too difficult for them (frustration, Csikszentmihalyi (1997)). On the other hand, the instructors/educators can provide affective/motivational support to the teams who are more bored than others. Such activities are either too easy for them (Csikszentmihalyi, 1997) or the team is not performing well (Kwon et al., 2014), or they are not interested in the activity at all.

Another approach would be to view children's dialogue as important for team interaction. For example, encouraging children into more playful talk (Pursi, Lipponen, & Sajaniemi, 2018) between them can be an answer to a situation of a negative emotion that persists over time. Overall, the most important aspect to consider is how to support children move on with their emotions during a coding activity. Naturally, many emotions appear in the learning process, and not only the positive ones are the ones to be valued. We need to support and keep the learners flow into the cycle of positive and negative emotions recognizing their value (Kort et al., 2001). This approach is supported by designing feedback interventions, which are messages that have been motivationally designed feedback messages (Narciss, 2008). We can contextualize this within the theories of self-regulated learning. The primary function of this feedback rests in guiding the learner to successfully regulate his or her learning process (Butler & Winne, 1995; Narciss, 2008). By letting the students understand their emotions, instructors can provide feedback that is considered a forward-facing mechanism with a fundamental orientation towards improving not only performance but also knowledge, identities, and values (Molloy, Noble, & Ajjawi, 2019).

5.3. Ethical considerations

While using data, such as videos captured from the camera and recognizing facial emotions/expressions, the children's privacy is considered. This becomes even more crucial when these data streams are used in interactive application and decision-making (e.g., detecting moments to support students/children). The context where facial data is used can be both engaging for the children and provoke disengagement (Sharma & Giannakos, 2021). When affective technologies (Bekele et al., 2013; Javed, Burns, Jeon, Howard, & Park, 2019; Okita, Ng-Thow-Hing, & Sarvadevabhatla, 2011) use emotions and/or facial expressions and emotions as a key factor in their protocols and/or analysis, they can have varied effects on how the children perceive their privacy if they are interacting with the technology in different settings, such as homes, schools, or outdoors. There have been efforts in CCI to raise cautions and suggest methods to deal with such concerns (Dowthwaite et al., 2020; Kawas et al., 2020; Papavlasopoulou et al., 2019; Van Mechelen, Baykal, Dindler, Eriksson, & Iversen, 2020).

One such recommendation is to inform the children about "how their data will be used?" so that contextual biases can be mitigated. For example, a short-term use, where the sensitive data is used to trigger the momentary adaptivity as a part of the interaction (e.g., task completion, frustration/confusion mitigation). A long-term use with proper and systematic anonymization processes creates more long-term effective and efficient routines (e.g., longitudinal engagement, skill improvement). These long-term routines can be properly and in-depth vetted by the teachers and parents, depending on the context of the study (e.g., homes, schools, outdoors). When it comes to the social aspects of the CCI studies using facial recognition and/or emotion recognition technologies, there are certain roles that the parents and/or teachers can play. These roles might increase children's acceptance of and engagement with the sensing technologies and help conduct the study smoothly and uninterruptedly (Sharma & Giannakos, 2021).

5.4. Limitations and future work

In this paper, we focus only on the affective states from one data source (i.e., the facial features). In recent times, with the advancements in the physiological sensors (Sharma, Papamitsiou, & Giannakos, 2019), and multimodal learning analytics (Giannakos, Sharma, Pappas, Kostakos, & Velloso, 2019), it has become easier to incorporate more modalities to understand other affective states. For example, stress and arousal including cognitive processes, such as attention, cognitive load and mental effort; and social processes such as dialogues. By doing this, we could gain a holistic understanding of the collaborative coding processes. Moreover, on the analytical front, this paper utilizes only Markovian analysis when it comes to the temporal analysis, which has certain disadvantages (Sharma, Papamitsiou, Olsen, & Giannakos, 2020). In the future, we will incorporate a longer history than the previous timestamp and move away from the Markov assumption to a more temporal analysis. Moreover, since in this paper data are presented as aggregate per team, in the future, we will investigate if there were specific individuals who had a major impact on the results of each team or acted as "influencers" (i.e., their emotions gradually affected the rest of the team). Further, this paper presents the relation between various variables in terms of correlations and regressions, in HCI, there is a call for the shift towards causality among the peers, and the different expressions (Pekmn et al., 2002). We will also explore the future causal relations between individual and joint emotions. Finally, some of the collaborative dynamics are not observed from the results. For example, which child is getting control of the situation and how/when this control passes to another peer. We also aim to address such collaborative learning/work-specific questions in future work.

6. Conclusion

In this contribution, we show that there are clear relationships between the different measurements of children's performance (i.e., objective and subjective) during coding activities and their emotions, measured from two different points of view (i.e., education-specific and control value theory-based). The results and implications show that it is important for researchers to consider a diversity of emotions while analyzing children's interaction among themselves in teams and with the technology. This diverse set of behavior would allow them to provide better support to the children than while using a limited set of emotions. From the results we can also show that there are different aspects of performance captured by the two sets of emotions, for example, the objective performance is better explained by the education-specific emotions and the negative valance emotions from the CVT-based emotions. On the other hand, subjective performance is better explained by the CVT-based emotions only. In the future, we would also consider the causal relations between these measurements to better support the design of feedback and reflection tools. Furthermore, the lack of any relationships between the education-specific emotions and subjective performance requires further investigation to achieve consensus from varied contexts.

Selection and participation of children

All the participants of the study were students from the Trondheim (Norway) region whose teachers have applied to participate in our workshops as an out-of-school activity. Studies took place at the university campus in specially designed rooms. Data related to the study were collected after permission from the National Data Protection Official for Research, following all the regulations and recommendations for research with children. A researcher contacted the teacher and the legal guardian of each child to get a written consent that permitted the data collection. The children were informed about the data collection process and their participation in the study was completely voluntary. They could withdraw their consent for the data collection at any time without affecting their participation in the coding activity.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Ainley, M., Corrigan, M., & Richardson, N. (2005). Students, tasks and emotions: Identifying the contribution of emotions to students' reading of popular culture and popular science texts. *Learning and Instruction*, *15*(5), 433–447.

Amos, B., Ludwiczuk, B., Satyanarayanan, M., et al. (2016). Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, *6*.

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.

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.

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). Openface: An open source facial behavior analysis toolkit. In *2016 IEEE winter conference on applications of computer vision* (pp. 1–10). IEEE, http://dx.doi.org/10.1109/WACV.2016.7477553.

Barnett, J., Coulson, M., & Foreman, N. (2010). Examining player anger in world of warcraft. In *Online worlds: Convergence of the real and the virtual* (pp. 147–160). Springer.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences*, *9*(4), 403–436.

Barrouillet, P. (2015). Theories of cognitive development: From Piaget to today. *Developmental Review*, *38*, 1–12.

Bekele, E., Zheng, Z., Swanson, A., Davidson, J., Warren, Z., & Sarkar, N. (2013). Virtual reality-based facial expressions understanding for teenagers with autism. In *International conference on universal access in human-computer interaction* (pp. 454–463). Springer.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. McGraw-Hill Book Company.

Bless, H. (2000). *The interplay of affect and cognition: The mediating role of general knowledge structures*. Cambridge University Press.

Bocconi, S., Chioccariello, A., Dettori, G., Ferrari, A., Engelhardt, K., Kampylis, P., et al. (2016). *68*, *Developing computational thinking in compulsory education: European Commission, JRC Science for Policy Report*.

Bosch, N., D'Mello, S., & Mills, C. (2013). What emotions do novices experience during their first computer programming learning session? In *International conference on artificial intelligence in education* (pp. 11–20). Springer.

Bowden, W. R. (2015). Collaboration, pedagogy, and media: Short-term summer programs emphasize project based and social emotional learning. *Journal of Media Literacy Education*, *7*(1), 72–76.

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the american educational research association, vol. 1* (p. 25).

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 1287–1294.

Brown, N. C., Sentance, S., Crick, T., & Humphreys, S. (2014). Restart: The resurgence of computer science in UK schools. *ACM Transactions on Computing Education (TOCE)*, *14*(2), 1–22.

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

Capdeferro, N., & Romero, M. (2012). Are online learners frustrated with collaborative learning experiences? *The International Review of Research in Open and Distributed Learning*, *13*(2), 26–44.

Carlsson, T., Winder, M., Eriksson, A. L., & Wallerstedt, S. M. (2020). Student characteristics associated with passing the exam in undergraduate pharmacology courses—A cross-sectional study in six university degree programs. *Medical Science Educator*, *30*(3), 1137–1144.

Chan, C. K. (2013). Towards a knowledge creation perspective. *The International Handbook of Collaborative Learning*, 437.

Chang, C.-K., & Tsai, Y.-T. (2018). Pair-programming curriculum development of motion-based game for enhancing computational thinking skills. In *2018 7th international congress on advanced applied informatics* (pp. 284–287). IEEE.

Chen, S.-J., & Caropreso, E. J. (2004). Influence of personality on online discussion. *Journal of Interactive Online Learning*, *3*(2), 1–17.

Cheng, M.-T., Huang, W.-Y., & Hsu, M.-E. (2020). Does emotion matter? An investigation into the relationship between emotions and science learning outcomes in a game-based learning environment. *British Journal of Educational Technology*, *51*(6), 2233–2251.

Clarebout, G., & Elen, J. (2001). The ParlEuNet-project: Problems with the validation of socio-constructivist design principles in ecological settings. *Computers in Human Behavior*, *17*(5–6), 453–464.

Csikszentmihalhi, M. (2020). *Finding Flow: The Psychology of Engagement with Everyday Life*. Hachette UK.

Csikszentmihalyi, M. (1997). Flow and the psychology of discovery and invention. In *Harper perennia, vol. 39*. New York.

Delaborde, A., Tahon, M., Barras, C., & Devillers, L. (2009). A Wizard-of-Oz game for collecting emotional audio data in a children-robot interaction. In *Proceedings of the international workshop on affective-aware virtual agents and social robots* (p. 5). ACM.

Denner, J. (2007). The girls creating games program: An innovative approach to integrating technology into middle school. *Meridian: A Middle School Computer Technologies Journal*, *1*(10).

Denner, J., Werner, L., Campe, S., & Ortiz, E. (2014). Pair programming: Under what conditions is it advantageous for middle school students? *Journal of Research on Technology in Education*, *46*(3), 277–296.

Denner, J., Werner, L., & Ortiz, E. (2012). Computer games created by middle school girls: Can they be used to measure understanding of computer science concepts? *Computers & Education*, *58*(1), 240–249.

Dewey, J. (2018). *Logic-the theory of inquiry*. Read Books Ltd.

Di Leo, I., Muis, K. R., Singh, C. A., & Psaradellis, C. (2019). Curiosity…confusion? Frustration! the role and sequencing of emotions during mathematics problem solving. *Contemporary Educational Psychology*, *58*, 121–137.

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*, *105*(4), 1082.

D'Mello, S., & Graesser, A. (2010). Modeling cognitive-affective dynamics with hidden Markov models. In *Proceedings of the annual meeting of the cognitive science society, vol. 32, no. 32*.

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

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

Dore, I. (2016). Shape-shifter and agenda setter: The influence of emotion in social work practice and practice education. *Social Work Education*, *35*(4), 469–481.

Dowd, J. E., Araujo, I., & Mazur, E. (2015). Making sense of confusion: Relating performance, confidence, and self-efficacy to expressions of confusion in an introductory physics class. *Physical Review Special Topics-Physics Education Research*, *11*(1), Article 010107.

Dowthwaite, L., Creswick, H., Portillo, V., Zhao, J., Patel, M., Vallejos, E. P., et al. (2020). " It's your private information. it's your life." young people's views of personal data use by online technologies. In *Proceedings of the interaction design and children conference* (pp. 121–134).

Eckmann, J.-P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *EPL (Europhysics Letters)*, *4*(9), 973.

Flook, L., Repetti, R. L., & Ullman, J. B. (2005). Classroom social experiences as predictors of academic performance. *Developmental Psychology*, *41*(2), 319.

Fong, K., & Kleiner, B. H. (2004). New development concerning the effect of work overload on employees. *Management Research News*.
Ford, D., & Parnin, C. (2015). Exploring causes of frustration for software developers. In *2015 IEEE/ACM 8th international workshop on cooperative and human aspects of software engineering* (pp. 115–116). IEEE.
Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology*, *2*(3), 300.
Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics—A ''hopeless'' issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *22*(4), 497.
Gachechiladze, D., Lanubile, F., Novielli, N., & Serebrenik, A. (2017). Anger and its direction in collaborative software development. In *2017 IEEE/ACM 39th international conference on software engineering: New ideas and emerging technologies results track* (pp. 11–14). IEEE.
Giannakos, M. N., Chorianopoulos, K., Inkpen, K., Du, H., & Johns, P. (2013). Understanding children's behavior in an asynchronous video-mediated communication environment. *Personal and Ubiquitous Computing*, *17*(8), 1621–1629.
Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, *48*, 108–119.
Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, *49*(2), 193–223.
Govender, D. W., & Govender, T. (2014). Using a collaborative learning technique as a pedagogic intervention for the effective teaching and learning of a programming course. *Mediterranean Journal of Social Sciences*, *5*(20), 1077.
Graziotin, D., Wang, X., & Abrahamsson, P. (2014). Software developers, moods, emotions, and performance. arXiv preprint arXiv:1405.4422.
Hager, P. E.-W. F.-J. (2002). *Facial action coding system. The manual on CD ROM*. Research Nexus Division of Network Information Research Corporation, Salt.
Hannay, J. E., Dybå, T., Arisholm, E., & Sjøberg, D. I. (2009). The effectiveness of pair programming: A meta-analysis. *Information and Software Technology*, *51*(7), 1110–1122.
Hanson, T. L., Austin, G., & Lee-Bayha, J. (2004). Ensuring that no child is left behind. How are student health risks & resilience related to the academic progress of schools? *WestEd*.
Harrington, N. (2005). It's too difficult! frustration intolerance beliefs and procrastination. *Personality and Individual Differences*, *39*(5), 873–883.
Harrold, N., Tan, C. T., Rosser, D., & Leong, T. W. (2014). CopyMe: An emotional development game for children. In *CHI'14 extended abstracts on human factors in computing systems* (pp. 503–506). ACM.
Heckel, C., & Ringeisen, T. (2017). Enjoyment and boredom in academic online-learning: Relations with appraisals and learning outcomes. *Stress and Anxiety: Coping and Resilience*, 127–136.
Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction*, *26*(1), 48–94.
Isohätälä, J., Näykki, P., & Järvelä, S. (2020). Cognitive and socio-emotional interaction in collaborative learning: Exploring fluctuations in students' participation. *Scandinavian Journal of Educational Research*, *64*(6), 831–851.
Israel, M., Wherfel, Q. M., Shehab, S., Ramos, E. A., Metzger, A., & Reese, G. C. (2016). Assessing collaborative computing: Development of the collaborative-computing observation instrument (C-COI). *Computer Science Education*, *26*(2–3), 208–233.
Javed, H., Burns, R., Jeon, M., Howard, A. M., & Park, C. H. (2019). A robotic framework to facilitate sensory experiences for children with autism spectrum disorder: A preliminary study. *ACM Transactions on Human-Robot Interaction (THRI)*, *9*(1), 1–26.
Jenkins, T. (2002). On the difficulty of learning to program. In *Proceedings of the 3rd annual conference of the LTSN centre for information and computer sciences, vol. 4, no. 2002* (pp. 53–58). Citeseer.
Jiménez, S., Juárez-Ramírez, R., Castillo, V. H., Licea, G., Ramírez-Noriega, A., & Inzunza, S. (2018). A feedback system to provide affective support to students. *Computer Applications in Engineering Education*, *26*(3), 473–483.
Jordan, M. E., & McDaniel Jr, R. R. (2014). Managing uncertainty during collaborative problem solving in elementary school teams: The role of peer influence in robotics engineering activity. *Journal of the Learning Sciences*, *23*(4), 490–536.
Judge, S., & Jahns, L. (2007). Association of overweight with academic performance and social and behavioral problems: An update from the early childhood longitudinal study. *Journal of School Health*, *77*(10), 672–678.
Kapoor, A., Mota, S., Picard, R. W., et al. (2001). Towards a learning companion that recognizes affect. In *AAAI fall symposium, vol. 543* (pp. 2–4).
Kapoor, A., & Picard, R. W. (2005). Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on multimedia* (pp. 677–682). ACM.
Kawas, S., Yuan, Y., DeWitt, A., Jin, Q., Kirchner, S., Bilger, A., et al. (2020). Another decade of IDC research: Examining and reflecting on values and ethics. In *Proceedings of the interaction design and children conference* (pp. 205–215).
Kinnunen, P., & Simon, B. (2010). Experiencing programming assignments in CS1: The emotional toll. In *Proceedings of the sixth international workshop on computing education research* (pp. 77–86).
Kinnunen, P., & Simon, B. (2012). My program is ok–am i? Computing freshmen's experiences of doing programming assignments. *Computer Science Education*, *22*(1), 1–28.
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.
Kostyuk, V., Almeda, M. V., & Baker, R. S. (2018). Correlating affect and behavior in reasoning mind with state test achievement. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 26–30).
Kruk, M. (2016). Investigating the changing nature of boredom in the english language classroom: Results of a study. In *Nowy Wymiar Filologii* (pp. 252–263). Wydawnictwo Naukowe Państwowej Wyższej Szkoły Zawodowej w Płocku Płock.
Kruk, M. (2021). Investigating the experience of boredom during reading sessions in the foreign language classroom. *Journal of Language & Education Volume*, *7*(3).
Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, *44*(1), 46–53.
Kwon, K., Hanrahan, A. R., & Kupzyk, K. A. (2017). Emotional expressivity and emotion regulation: Relation to academic functioning among elementary school children. *School Psychology Quarterly*, *32*(1), 75.
Kwon, K., Liu, Y.-H., & Johnson, L. P. (2014). Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers & Education*, *78*, 185–200.
Lajoie, S. P., Lee, L., Poitras, E., Bassiri, M., Kazemitabar, M., Cruz-Panesso, I., et al. (2015). The role of regulation in medical student learning in small groups: Regulating oneself and others' learning and emotions. *Computers in Human Behavior*, *52*, 601–616.
Lehman, B., D'Mello, S., Strain, A., Mills, C., Gross, M., Dobbins, A., et al. (2013). Inducing and tracking confusion with contradictions during complex learning. *International Journal of Artificial Intelligence in Education*, *22*(1–2), 85–105.
Leite, I., Castellano, G., Pereira, A., Martinho, C., Paiva, A., & McOwan, P. W. (2009). Designing a game companion for long-term social interaction. In *Proceedings of the international workshop on affective-aware virtual agents and social robots* (p. 10). ACM.
Leite, I., Henriques, R., Martinho, C., & Paiva, A. (2013). Sensors in the wild: Exploring electrodermal activity in child-robot interaction. In *Proceedings of the 8th ACM/IEEE international conference on human-robot interaction* (pp. 41–48). IEEE Press.
Lescano, G., & Costaguta, R. (2018). COLLAB: Conflicts and sentiments in chats. In *Proceedings of the XIX international conference on human computer interaction* (p. 33). ACM.
Linnenbrink, E. A., & Pintrich, P. R. (2002). The role of motivational beliefs in conceptual change. In *Reconsidering conceptual change: Issues in theory and practice* (pp. 115–135). Springer.
Linnenbrink, E., & Pintrich, P. (2003). Motivation, affect, and cognitive processing: What role does affect play. In *Annual meeting of the American educational research association*. Chicago, IL.
Linnenbrink-Garcia, L., Rogat, T. K., & Koskey, K. L. (2011). Affect and engagement during small group instruction. *Contemporary Educational Psychology*, *36*(1), 13–24.
Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. (2013). Sequences of frustration and confusion, and learning. In *Educational data mining 2013*. Citeseer.
Loderer, K., Pekrun, R., & Lester, J. C. (2018). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, Article 101162.
Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, *70*, Article 101162.
Makewa, L. N., Gitonga, D., Ngussa, B., Njoroge, S., & Kuboja, J. (2014). Frustration factor in group collaborative learning experiences. *American Journal of Educational Research*, *2*(11A), 16–22.
McDaniel, B., D'Mello, S., King, B., Chipman, P., Tapp, K., & Graesser, A. (2007). Facial features for affective state detection in learning environments. In *Proceedings of the annual meeting of the cognitive science society, vol. 29, no. 29*.
McDowell, C., Werner, L., Bullock, H., & Fernald, J. (2002). The effects of pair-programming on performance in an introductory programming course. In *Proceedings of the 33rd SIGCSE technical symposium on computer science education* (pp. 38–42).
McDowell, C., Werner, L., Bullock, H. E., & Fernald, J. (2003). The impact of pair programming on student performance, perception and persistence. In *25th international conference on software engineering, 2003. Proceedings* (pp. 602–607). IEEE.

McQuiggan, S. W., & Lester, J. C. (2007). Modeling and evaluating empathy in embodied companion agents. *International Journal of Human-Computer Studies*, *65*(4), 348–360.
Meyer, D. K., & Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist*, *37*(2), 107–114.
Mitra, R., & Chavan, P. (2019). DEBE feedback for large lecture classroom analytics. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 426–430).
Molloy, E., Noble, C., & Ajjawi, R. (2019). Attending to emotion in feedback. In *The impact of feedback in higher education* (pp. 83–105). Springer.
Moreno-León, J., Robles, G., & Román-González, M. (2015). Dr. Scratch: Automatic analysis of scratch projects to assess and foster computational thinking. *RED. Revista De Educación a Distancia*, (46), 1–23.
Moreno-León, J., Román-González, M., Harteveld, C., & Robles, G. (2017). On the automatic assessment of computational thinking skills: A comparison with human experts. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems* (pp. 2788–2795).
Narciss, S. (2008). Feedback strategies for interactive learning tasks. In *Handbook of research on educational communications and technology, vol. 3* (pp. 125–144).
O'Donnell, C., Buckley, J., Mahdi, A., Nelson, J., & English, M. (2015). Evaluating pair-programming for non-computer science major students. In *Proceedings of the 46th ACM technical symposium on computer science education* (pp. 569–574).
Okita, S. Y., Ng-Thow-Hing, V., & Sarvadevabhatla, R. K. (2011). Multimodal approach to affective human-robot interaction design with children. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *1*(1), 1–29.
Panitz, T. (1999). *The case for student centered instruction via collaborative learning paradigms*. ERIC.
Papavlasopoulou, S. (2019). *Making-based coding activities for young students: Design meaningful learning experiences*. NTNU.
Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2017). Empirical studies on the maker movement, a promising approach to learning: A literature review. *Entertainment Computing*, *18*, 57–78.
Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2019). Exploring children's learning experience in constructionism-based coding activities through design-based research. *Computers in Human Behavior*, *99*, 415–427.
Papavlasopoulou, S., Sharma, K., & Giannakos, M. N. (2018). How do you feel about learning to code? Investigating the effect of children's attitudes towards coding using eye-tracking. *International Journal of Child-Computer Interaction*.
Papert, S. (1990). *Children, computers and powerful ideas*. New York: Basic Books.
Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.
Pekmn, R., Goetz, T., Titz, W., et al. (2002). Academic emotions in students'' self regulated learning and achievement: A program of quantitative and qualitative research. *Educational Psychologist*, *37*, 91–106.
Pekrun, R. (2006). c. *Educational Psychology Review*, *18*(4), 315–341. http://dx.doi.org/10.1007/s10648-006-9029-9.
Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In *Emotion in education* (pp. 13–36). Elsevier.
Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48.
Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002a). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105.
Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002b). Positive emotions in education.
Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, *88*(5), 1653–1670.
Perry, D., & Aragon, C. (2012). Measuring distributed affect in collaborative games. In *Proceedings of the ACM 2012 conference on computer supported cooperative work companion* (pp. 195–198). ACM.
Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., et al. (2004). Affective learning—A manifesto. *BT Technology Journal*, *22*(4), 253–269.
Porter, L., & Simon, B. (2013). Retaining nearly one-third more majors with a trio of instructional best practices in CS1. In *Proceeding of the 44th ACM technical symposium on computer science education* (pp. 165–170).
Pursi, A., Lipponen, L., & Sajaniemi, N. K. (2018). Emotional and playful stance taking in joint play between adults and very young children. *Learning, Culture and Social Interaction*, *18*, 28–45.
Putwain, D. W., Becker, S., Symes, W., & Pekrun, R. (2018). Reciprocal relations between students' academic enjoyment, boredom, and achievement over time. *Learning and Instruction*, *54*, 73–81.
Putwain, D. W., Pekrun, R., Nicholson, L. J., Symes, W., Becker, S., & Marsh, H. W. (2018). Control-value appraisals, enjoyment, and boredom in mathematics: A longitudinal latent interaction analysis. *American Educational Research Journal*, *55*(6), 1339–1368.
Radermacher, A. D., & Walia, G. S. (2011). Investigating the effective implementation of pair programming: An empirical investigation. In *Proceedings of the 42nd ACM technical symposium on computer science education* (pp. 655–660).
Reis, R. C. D., Isotani, S., Rodriguez, C. L., Lyra, K. T., Jaques, P. A., & Bittencourt, I. I. (2018). Affective states in computer-supported collaborative learning: Studying the past to drive the future. *Computers & Education*, *120*, 29–50.
Richey, J. E., Andres-Bray, J. M. L., Mogessie, M., Scruggs, R., Andres, J. M., Star, J. R., et al. (2019). More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education*, *139*, 173–190.
Rodrigo, M. M. T., Baker, R. S., Jadud, M. C., Amarra, A. C. M., Dy, T., Espejo-Lahoz, M. B. V., et al. (2009). Affective and behavioral predictors of novice programmer achievement. In *Proceedings of the 14th annual ACM SIGCSE conference on innovation and technology in computer science education* (pp. 156–160).
Rodrigo, M., Mercedes, T., d Baker, R. S., McLaren, B. M., Jayme, A., & Dy, T. T. (2012). Development of a workbench to address the educational data mining bottleneck. *International Educational Data Mining Society*.
Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The Journal of the Learning Sciences*, *2*(3), 235–276.
Royston, J. (1982). Algorithm AS 181: The W test for normality. *Applied Statistics*, 176–180.
Salleh, N., Mendes, E., & Grundy, J. (2010). Empirical studies of pair programming for CS/SE teaching in higher education: A systematic literature review. *IEEE Transactions on Software Engineering*, *37*(4), 509–525.
Schermuly, C. C. (2014). Negative effects of coaching for coaches: An explorative study. *International Coaching Psychology Review*, *9*(2), 165–180.
Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Broda, M., Spicer, J., et al. (2016). Investigating optimal learning moments in US and finnish science classes. *Journal of Research in Science Teaching*, *53*(3), 400–421.
Schultz, J. L., Wilson, J. R., & Hess, K. C. (2010). Team-based classroom pedagogy reframed: The student perspective. *American Journal of Business Education*, *3*(7), 17–24.
Seegers, G., & Boekaerts, M. (1993). Task motivation and mathematics achievement in actual task situations. *Learning and Instruction*, *3*(2), 133–150.
Seyam, M., & McCrickard, D. S. (2016). Teaching mobile development with pair programming. In *Proceedings of the 47th ACM technical symposium on computing science education* (pp. 96–101).
Shahid, S., Krahmer, E., Swerts, M., & Mubin, O. (2010). Child-robot interaction during collaborative game play: Effects of age and gender on emotion and experience. In *Proceedings of the 22nd conference of the computer-human interaction special interest group of Australia on computer-human interaction* (pp. 332–335). ACM.
Sharma, K., & Giannakos, M. (2021). Sensing technologies and child–computer interaction: Opportunities, challenges and ethical considerations. *International Journal of Child-Computer Interaction*, *30*, Article 100331.
Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology*.
Sharma, K., Papamitsiou, Z., Olsen, J. K., & Giannakos, M. (2020). Predicting learners' effortful behaviour in adaptive assessment using multimodal data. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 480–489).
Sharma, K., Papavlasopoulou, S., & Giannakos, M. (2019). Joint emotional state of children and perceived collaborative experience in coding activities. In *Proceedings of the 18th ACM international conference on interaction design and children* (pp. 133–145). ACM.
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.
Simonton, K. L., & Garn, A. (2019). Exploring achievement emotions in physical education: The potential for the control-value theory of achievement emotions. *Quest*, *71*(4), 434–446.
Simonton, K. L., & Garn, A. C. (2020). Control–value theory of achievement emotions: A closer look at student value appraisals and enjoyment. *Learning and Individual Differences*, *81*, Article 101910.
So, H.-J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, *51*(1), 318–336.
Spaulding, S., Gordon, G., & Breazeal, C. (2016). Affect-aware student models for robot tutors. In *Proceedings of the 2016 international conference on autonomous agents & multiagent systems* (pp. 864–872). International Foundation for Autonomous Agents and Multiagent Systems.
Sridhar, P. K., Chan, S. W., & Nanayakkara, S. (2018). Going beyond performance scores: Understanding cognitive-affective states in kindergarteners. In *Proceedings of the 17th ACM conference on interaction design and children* (pp. 253–265). ACM.

Stahl, G., Law, N., Cress, U., & Ludvigsen, S. (2014). Analyzing roles of individuals in small-group collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, *9*(4), 365–370.
Sullivan, F. R., & Wilson, N. C. (2015). Playful talk: Negotiating opportunities to learn in collaborative groups. *Journal of the Learning Sciences*, *24*(1), 5–52.
Suzuki, K. (2015). Social imaging technology to identify and represent social behaviors. In *Adjunct proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2015 ACM international symposium on wearable computers* (pp. 907–908). ACM.
Takano, Y., & Suzuki, K. (2014). Affective communication aid using wearable devices based on biosignals. In *Proceedings of the 2014 conference on interaction design and children* (pp. 213–216). ACM.
Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. *Computers As Cognitive Tools*, 229–258.
Tisza, G., & Markopoulos, P. (2021). Understanding the role of fun in learning to code. *International Journal of Child-Computer Interaction*, *28*, Article 100270.
Troiano, G. M., Snodgrass, S., Argımak, E., Robles, G., Smith, G., Cassidy, M., et al. (2019). Is my game OK Dr. Scratch? Exploring programming and computational thinking development via metrics in student-designed serious games for STEM. In *Proceedings of the 18th ACM international conference on interaction design and children* (pp. 208–219).
Tsai, T.-W., Lo, H. Y., & Chen, K.-S. (2012). An affective computing approach to develop the game-based adaptive learning material for the elementary students. In *Proceedings of the 2012 joint international conference on human-centered computer environments* (pp. 8–13). ACM.
Tsan, J., Lynch, C. F., & Boyer, K. E. (2018). ''Alright, what do we need?'': A study of young coders' collaborative dialogue. *International Journal of Child-Computer Interaction*, *17*, 61–71.
Tuhkala, A., Wagner, M.-L., Nielsen, N., Iversen, O. S., & Kärkkäinen, T. (2018). Technology comprehension: Scaling making into a national discipline. In *Proceedings of the conference on creativity and making in education* (pp. 72–80).
Tze, V., Parker, P., & Sukovieff, A. (2022). Control-Value theory of achievement emotions and its relevance to school psychology. *Canadian Journal of School Psychology*, *37*(1), 23–39.
Umapathy, K., & Ritzhaupt, A. D. (2017). A meta-analysis of pair-programming in computer programming courses: Implications for educational practice. *ACM Transactions on Computing Education (TOCE)*, *17*(4), 1–13.
Van Diggelen, W., & Overdijk, M. (2007). Small-group face-to-face discussions in the classroom: a new direction of CSCL research. In *Proceedings of the 8th iternational conference on computer supported collaborative learning* (pp. 727–736). International Society of the Learning Sciences.
Van Mechelen, M., Baykal, G. E., Dindler, C., Eriksson, E., & Iversen, O. S. (2020). 18 Years of ethics in child-computer interaction research: A systematic literature review. In *Proceedings of the interaction design and children conference* (pp. 161–183).
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.
Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes (E. Rice, Ed. & Trans)*. Cambridge, MA: Harvard University Press.(Original work published 1930, 1933.
Werner, L., & Denning, J. (2009). Pair programming in middle school: What does it look like? *Journal of Research on Technology in Education*, *42*(1), 29–49.
Werner, L. L., Hanks, B., & McDowell, C. (2004). Pair-programming helps female computer science students. *Journal on Educational Resources in Computing (JERIC)*, *4*(1), 4–es.
Werner, J. M., & Lester, S. W. (2001). Applying a team effectiveness framework to the performance of student case teams. *Human Resource Development Quarterly*, *12*(4), 385–402.
Williams, L., & Upchurch, R. L. (2001). In support of student pair-programming. *ACM SIGCSE Bulletin*, *33*(1), 327–331.
Wu, C.-H., Huang, Y.-M., & Hwang, J.-P. (2016). Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology*, *47*(6), 1304–1323.
Yip, W., Wishart, D., & Barrett, D. 2020. Safe and unsafe drivers.
Zhang, L. (2008). Metaphorical affect sensing in an intelligent conversational agent. In *Proceedings of the 2008 international conference on advances in computer entertainment technology* (pp. 100–106). ACM.
Zhang, F., Markopoulos, P., & Bekker, T. (2020). Children's emotions in design-based learning: A systematic review. *Journal of Science Education and Technology*, *29*(4), 459–481.
Zhang, F., Markopoulos, P., Bekker, T., Schüll, M., & Paule-Ruíz, M. (2019). EmoForm: Capturing children's emotions during design based learning. In *Proceedings of FabLearn 2019* (pp. 18–25).
Zhou, Q., Main, A., & Wang, Y. (2010). The relations of temperamental effortful control and anger/frustration to Chinese children's academic achievement and social adjustment: A longitudinal study. *Journal of Educational Psychology*, *102*(1), 180.