Evaluating emotion visualizations using AffectVis, an affect-aware dashboard for students

Leony Derick
Department of Telematic Engineering, Universidad Carlos III de Madrid, Madrid, Spain

Gayane Sedrakyan
Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium

Pedro J. Munoz-Merino and Carlos Delgado Kloos
Department of Telematic Engineering, Universidad Carlos III de Madrid, Madrid, Spain, and

Katrien Verbert
Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract

Purpose – The purpose of this paper is to evaluate four visualizations that represent affective states of students.
Design/methodology/approach – An empirical-experimental study approach was used to assess the usability of affective state visualizations in a learning context. The first study was conducted with students who had knowledge of visualization techniques ($n = 10$). The insights from this pilot study were used to improve the interpretability and ease of use of the visualizations. The second study was conducted with the improved visualizations with students who had no or limited knowledge of visualization techniques ($n = 105$).
Findings – The results indicate that usability, measured by perceived usefulness and insight, is overall acceptable. However, the findings also suggest that interpretability of some visualizations, in terms of the capability to support emotional awareness, still needs to be improved. The level of students' awareness of their emotions during learning activities based on the visualization interpretation varied depending on previous knowledge of information visualization techniques. Awareness was found to be high for the most frequently experienced emotions and activities that were the most frustrating, but lower for more complex insights such as interpreting differences with peers. Furthermore, simpler visualizations resulted in better outcomes than more complex techniques.
Originality/value – Detection of affective states of students and visualizations of these states in computer-based learning environments have been proposed to support student awareness and improve learning. However, the evaluation of visualizations of these affective states with students to support awareness in real life settings is an open issue.
Keywords Learning analytics, Interactive learning environments, Human-computer interface, Visualization evaluation
Paper type Case study

1. Introduction

Emotions are known to play an important role in learning (Kort *et al.*, 2001; Trigwell *et al.*, 2012). Emotions drive attention, which, in turn, drives learning and memory (Värlander, 2008). Emotions are often a more powerful determinant of our behavior than our brain's logical and rational processes (Sylwester, 1994). Furthermore, emotions play an essential role in studies on attitudes and motivation (Pintrich, 2003; Meyer and Turner, 2002). Several studies found that students experience a rich diversity of both positive and negative emotions in academic settings (Pekrun *et al.*, 2002). Prior research has highlighted the importance of supporting learner awareness of these emotions (Ashkanasy and Dasborough, 2003). Information on affective states can, for instance, help students (or stimulate interest) to reflect on the type of emotions they felt, the activities that generated certain emotions or their evolution over time. By analyzing such information, students can take a pro-active role in regulating their learning as well as taking decisions on their improvement needs during learning processes, based for instance on information from studies that relate learning outcomes with affective states (Baker *et al.*, 2010).

Recent research shows increased interest in the automatic detection of emotions in various contexts. There are studies that propose different methodologies and detectors of emotions that also demonstrate practical applications in different learning contexts. Examples include emotion detection algorithms based on facial and gesture recognition (Burleson, 2006). Such algorithms are mostly based on human body signals, such as brainwaves captured with various sensors (Azcarraga *et al.*, 2014). Several studies attempted to correlate such data with student actions in different learning environments such as Intelligent Tutoring Systems (Pardos *et al.*, 2013), MOOCs (Leony *et al.*, 2015), or course-specific environments (Leony *et al.*, 2013a). Recent research has also shown interest in biofeedback based on analysis of multi-modal data collected from various wearable sensors during learning tasks. In some studies, information on emotions is processed based on manual input by students when interacting with a learning environment (Muñoz-Merino *et al.*, 2014).

An important issue is that information about affective states, such as the type and intensity of the experienced emotion, should be presented in computer-based learning environments in an intuitive way to the different stakeholders, including teachers, students, and managers. Visualization techniques are one of the most used techniques to present such information in the context of so-called learning dashboards (Verbert *et al.*, 2014). The goal is to support stakeholders to gain insight from these visualizations, i.e. to provide information that can be of utility to support awareness, reflection, and decision making (Verbert *et al.*, 2013). There are some works that present visualizations of emotions in computer-based learning environments (Leony *et al.*, 2013b). However, to our knowledge, no studies can be found that evaluate the capabilities of different visualization techniques to support awareness of emotions in learning environments. Also, most evaluations of learning visualizations are done by teachers (Verbert *et al.*, 2014). Empirical studies focusing on the evaluation of visualizations by students that provide insight into the usability of these visualizations are largely lacking.

In this manuscript, we focus specifically on evaluating the usability of visualizations of affective states with students using well-known usability assessment constructs such as the perceived usefulness and insight. Usability, also, refers to the ease of use (Davis, 1989) of the visualizations. We define the perceived usefulness as the perception of students about the importance of each one of the visualizations for the learning process. Insight is defined as the extent to which students can interpret the presented visualizations in a correct way (North, 2006). This manuscript aims to present the first evaluations and experiences from the use of visualizations of affective states for students

and attempts to identify future research directions in this domain. The research questions are the following:

RQ1. How usable are the visualizations that we have developed for students in terms of their perceived usefulness and ease of use?

RQ2. Which insights are supported by the affective visualizations for students regarding their capability to support awareness?

In this work, we assess the usability of different visualizations using different groups of students in higher education. Bachelor and Master level students from two different study programs at two universities participated in the studies. The pilot study was conducted with a group of students with a background in visualization techniques, whereas the second study was conducted with a group of students with little knowledge of visualization techniques. The insights from the first user study ($n = 10$) were used to improve the visualizations. We used suggestions of these students to create additional visualizations. The second user study was conducted with the enhanced environment with a larger group of students ($n = 105$).

The rest of the paper is organized as follows: Section 2 presents related work on visualizations of affective states in the context of learning dashboards. Section 3 presents our methodology. Section 4 describes the AffectVis dashboard, including four different visualizations of affective states. Sections 5 and 6 present two user studies conducted in the context of two different courses, detailing the participants, data collection, data analysis, and post-study interview results. Section 7 discusses the findings and limitations of the work. Finally, Section 8 concludes the work proposing some future research directions.

2. Background
2.1 Learning analytics
Different dimensions have to be considered when developing learning analytics applications (Greller and Drachsler, 2012), including internal limitations, external constraints, instruments, data, objectives, and stakeholders. The objectives can be twofold: reflection and prediction. In this study, we focus on students as stakeholders and reflection as an objective.

Instruments usually rely on either information retrieval or information visualization technologies, or a combination of both. Information retrieval intends to infer high-level information from the analysis of raw data. Examples of this high-level information can be student characteristics, such as learning behavior patterns, and future performance indicators (Muñoz-Merino *et al.*, 2013). These indicators are then visualized to provide useful insights for teachers, students, and managers (Ruipérez-Valiente *et al.*, 2014). Visualizations are known to support self-regulated actions of learners in online environments, as they can simulate social engagement and reflection appropriate for the context of the learner (Glahn, 2009).

Detection of affective states in educational settings has been explored previously by several researchers in the field (Baker *et al.*, 2010; Jaques and Vicari, 2007; Burleson, 2006; Azcarraga *et al.*, 2014; Pardos *et al.*, 2013). Leony *et al.* (2013a) presented a concrete case of inference of emotions from interaction data in a programming environment. The approach consists of a set of Hidden Markov Models (HMMs). During a programming task, students were asked to provide information about their affective state. This information was used to train the HMMs that were later used to predict emotions. In another approach, Leony *et al.* (2015) used a rule-based model for each emotion of interest, contextualizing emotion detection in MOOCs. In this work, frustration is for instance understood to occur when students either frequently fail exercises or fail an exercise about a topic that they thought they had sufficient knowledge about.

In this paper, we will focus specifically on visualizing data about affective states of learners and evaluating the usability measured by usefulness and insight of visualizations. Several studies have explored the typology of affective states that occur during learning (DMello, 2012). The results of previous research in this domain indicate that the basic emotions identified by Friesen and Ekman (1978), such as anger, fear, sadness, joy, disgust, and surprise, typically do not play a significant role in learning (Kort *et al.*, 2001). Several studies have also identified subsets of affective states that typically do play a significant role in learning, at least in the case of college students. Craig *et al.* (2004) for instance found evidence for a link between learning and the affective states of confusion, flow, and boredom. D'Mello *et al.* (2006) in addition found significant relationships for happiness (Eureka), confusion, and frustration, but not for boredom. In this manuscript, we make use of the most common subset of affective states based on prior research suggestions, namely: frustration, confusion, boredom, happiness, and motivation. We visualize information on these five affective states experienced by students during their learning activities in a learning dashboard for students. We present the results of two user studies that assess the usability of different visualization techniques for these affective states.

2.2 Learning dashboards

Dashboards are instruments intended to improve decision making by amplifying or directing cognition and capitalizing on human perceptual capabilities (Yigitbasioglu and Velcu, 2012). In a learning context, dashboards aim to support learning process awareness, ultimately targeting regulation of learning (Sedrakyan, Järvelä and Kirschner, 2016; Sedrakyan *et al.*, 2017; Sedrakyan, Malmberg, Verbert, Järvelä and Kirschner, 2018). In recent years, several dashboard applications have been developed to support learning or teaching. Such dashboards provide graphical representations of the current and historical state of a learner to support decision making (Few, 2006). Dashboards have been deployed to support learners or teachers, or both, and used in traditional face-to-face, group work, or online/blended learning (Verbert *et al.*, 2014). Examples of dashboards that are used to support face-to-face teaching include Backstage (Pohl *et al.*, 2012), Classroom Salon (Barr and Gunawardena, 2012), and Participation Tool (Janssen *et al.*, 2007). The overall objective of these dashboards is to stimulate learner engagement during face-to-face sessions. Several dashboards also focus on group work and collaboration. TinkerLamp (Son, 2012) and Collaid (Maldonado *et al.*, 2012) are some prominent examples. Most dashboards, however, focus on online or blended learning. Course signals (Arnold and Pistilli, 2012) is one of the more prominent examples in this category. The dashboard predicts and visualizes learning outcomes based on three data sources: grades in the course so far, time on task, and past performance. Most of our work is also part of this category.

In the context of learning dashboards, there are a few interesting observations that are relevant to the content of this paper:

- One observation is that usability evaluations have been conducted most often with teachers. Teachers were often asked to indicate how useful they think a dashboard would be for learners. Such a perceived usefulness evaluation was conducted for instance with both student inspector (Scheuer and Zinn, 2007) and LOCO-Analyst (Ali *et al.*, 2012), both yielding positive results. Results of our evaluations with SAM (Govaerts *et al.*, 2012) and StepUp! (Santos *et al.*, 2013) indicate that the perceived usefulness is often higher for teachers than for students. In this paper, we focus specifically on evaluations with students. In contrast to earlier studies, which were often conducted with a relatively small number of participants, we present results of a case study with a relatively large number of students.

- Also, most dashboards focus on visualizations of utilized resources, time, test results, and social interactions (Verbert *et al.*, 2014). To the best of our knowledge, only a few dashboards have been presented that focus on the representation of student emotions. In a recent study, Ruiz *et al.* (2016) focus on the methodological aspects of developing dashboards that support emotion-related information. Only one study was conducted that evaluates the utility aspects of inclusion of emotion-related information into learning dashboards (GhasemAghaei *et al.*, 2016). The focus of the study is the utility of such a dashboard for instructors. In our work, we focus on visualization of affective data for learners, motivated by the fact that such data have shown to be an important player in learning behavior regulation (Baker *et al.*, 2010).

- Finally, little is known about the effectiveness of different visualization techniques to give students insight into their learning-related data. Different visualizations have been proposed in earlier work, but to which extent these visualizations can be interpreted in correctly by students, and which techniques work better than others, both need further research.

3. General methodology

In this work, we followed the principles of Design Science in information systems research, which targets building and evaluating innovative artifacts to help understand and solve knowledge problems (Von Alan *et al.*, 2004). Our artifact includes a learning dashboard that shows emotion-related data during a learning process. The goal of the dashboard is to support student awareness on their affective states during their learning activities.

We use visualization techniques to represent affective states in the context of a learning dashboard, motivated by the fact that interactive visualization techniques are known to support effective understanding of data, reasoning, and decision making (Keim, 2002). Several visualizations have been developed with the goal to allow students to obtain insight into affective information. We designed, implemented and deployed visualizations of a relevant subset of emotions based on prior studies, as explained in the previous section. These visualizations represent the intensity levels of emotions, learning activities during which the emotions were experienced by students, the evolution of emotions over time, as well as comparisons with data of peers.

An empirical, experimental study approach was used to assess the usability of such visualizations deployed in a dashboard (the design artifact). Two experiments have been conducted in two different universities with two different groups of students: students with and students without previous knowledge of information visualization techniques. During the experiments, students completed different learning tasks, further referred to as sessions. At the end of each session, students provided information about their emotions by answering a set of basic questions in an online survey, such as "indicate how frequently you felt motivated/happy/confused/frustrated/bored during this learning activity." Based on the input of students, the learning dashboard generated affective visualizations. Context information about the students was also collected: students completed a questionnaire about their personal characteristics, such as gender, age, and previous knowledge on information visualization techniques.

The dashboard with different visualizations of affective states was deployed and provided to students. The following data measurements were used:

- a five-position Likert-type scale was used to score subjective judgments of students about the proposed affective visualization method, such as ease of use, perceived utility, and insight;

- the System Usability Scale (SUS) method (Brooke, 1996) has been used to measure the usability;

- insight was measured using comparison between the actual data and student perceptions: exploratory correlation analyses have been performed to study the differences of students' visualizations interpretations with the intended goal of the visualization; and

- subjective perceptions and insight have been, in addition, explored using a set of objective questions in a post-study interview.

To isolate the impact of pro-social behavior (Mitchell and Jolley, 2012), the anonymity of participants was ensured by not disclosing any identifiable information.

4. AffectVis: a visual learning dashboard of affective states and learning activities in projects

For our studies, we have developed four visualizations with the general objective of allowing learners to reflect on their affective states and their connection with specific learning activities. The visualizations are web based. Thus, the only tool needed to access them is a web browser with JavaScript capabilities.

Figure 1 shows the first visualization (radial visualization), which includes an improved version of the visualization presented by Leony, Parada, Muñoz-Merino, Pardo and Delgado Kloos (2013). The technique makes use of a set of polar bars to present the average frequency of each affective state experienced per each learning activity. Affective states of a
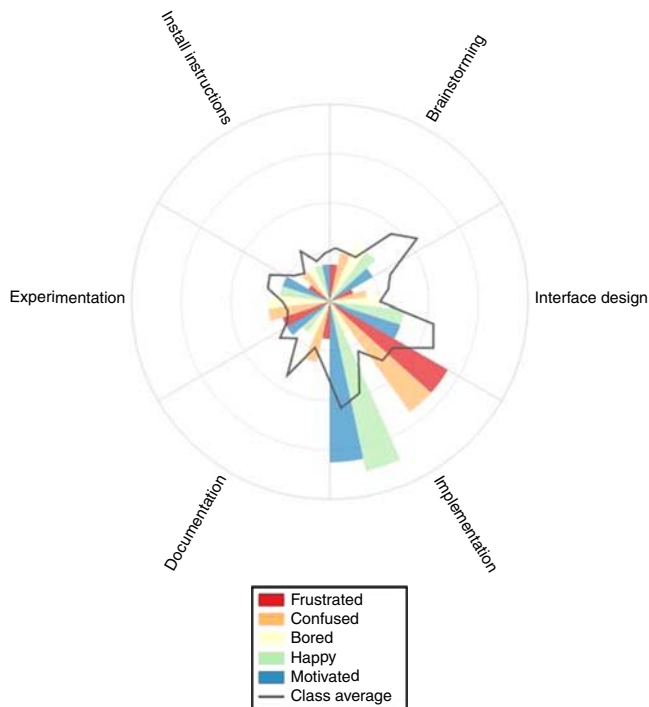

Figure 1.
The radial visualization is showing the frequency of each affective state for each activity

Source: Polar bars show the values for the active student, while the solid line shows the class average. Sedrakyan *et al.* (2017) (used with permission)

learner are differentiated through the color of the bar, while labels are used to represent the associated activity. The solid line shows the average value of the class for each emotion and activity.

The second visualization (timeline visualization) presents the evolution of time dedication of each student during the course, as well as the average time dedication and emotion evolutions of the whole class. The visualization represents the accumulated time dedication of students: when the student selects a point of time on the horizontal axis, the values of the vertical axis indicate the accumulated levels of time dedicated to learning activities during the course until that moment. In addition, the timeline visualizes the evolution of each emotion during the course. Figure 2 presents an example of the timeline visualization used in the scope of this work.

The third visualization is a heatmap visualization, in which columns represent time units, such as days, weeks, and months, and rows represent students. Each affective dimension is represented by a cell, while the frequency level of each emotion is represented through the intensity of the cell color (a more intense color represents a higher level of the emotion). A portion of this visualization is shown in Figure 3.

Lastly, we designed a scatterplot visualization. In this visualization, each affective dimension has a different scatterplot associated to it. The X-axis corresponds to the exact date and time when the emotion takes place, and the Y-axis presents the frequency value of the emotion. Bubble sizes represent the amount of work dedication, and bubble colors indicate whether the data point belongs to the active student (blue) or a peer. Figure 4 presents an example scatterplot for the emotion "confusion."

In its current form, the visualizations in the AffectVis dashboard rely on the data collected through systematic surveys of students about the typology and intensity of their emotion per different learning activity. Further details are described in the next sections.

5. Pilot study with a small group of students with knowledge in information visualization: user study 1

The main purpose of this user study was to perform an initial exploratory analysis of the developed visualizations with a small number of students. The radial visualization and the timeline visualization were deployed and evaluated in this study. Based on feedback and input from students participating in this study, the visualizations were improved. Moreover, two additional visualizations were developed and deployed for the second, more elaborate, user study.
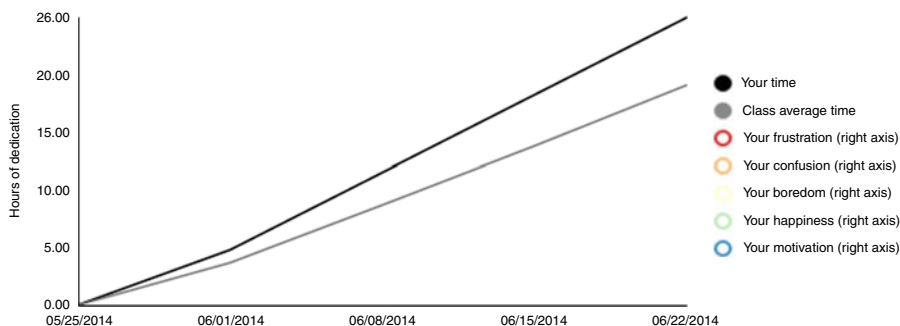
Source: Students can also see the time dedication average of the class (Sedrakyan *et al*., 2017) (used with permission)

Figure 2.
Timeline visualization
of the accumulated
amount of time
dedication and
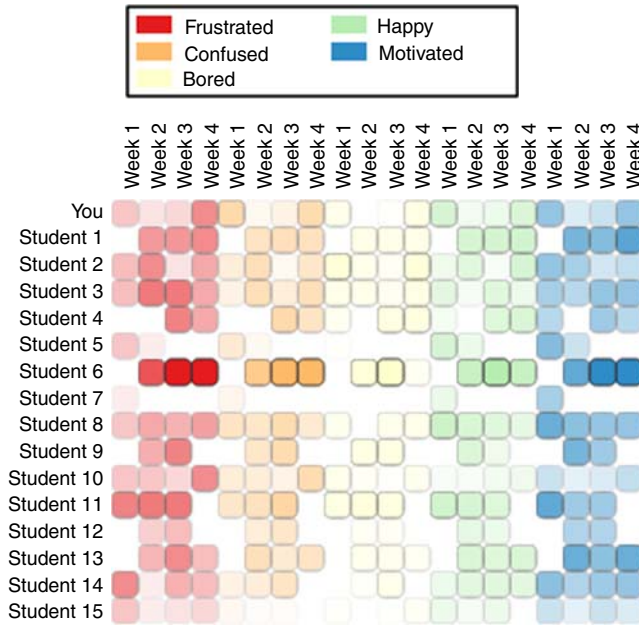the frequency of
affective dimensions

Figure 3.
Heatmap visualization of emotion frequency for each learner and each week

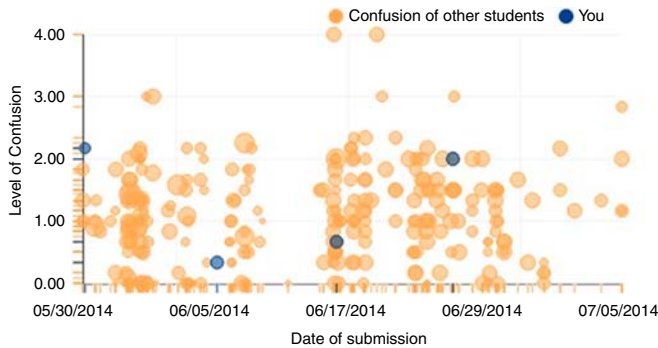**Source:** Sedrakyan *et al.* (2017) (used with permission)

Figure 4.
Scatterplot visualization showing the relation between the emotion frequency and work dedication along time

Source: Sedrakyan *et al.* (2017) (used with permission)

The first study was conducted with Master level students at Vrije Universiteit Brussel in Belgium. The profile of students, having knowledge about visualizations, was beneficial for obtaining targeted feedback. This user study would also allow observing differences between students with knowledge on visualizations and students without such knowledge (the students of user study 2). As indicated above, the radial visualization and the timeline were evaluated in this pilot study. At this stage, the timeline visualization included the aggregated time dedication of the student to different learning activities and the average time dedication of his/her peers.

The user study was conducted in the context of the course project, which lasted five weeks, from late February to early April of 2014.

5.1 Demographics of participants in user study 1

This user study was conducted in the context of a course on information visualization at a graduate level (Master degree program). First, students received theoretical and practical sessions about different concepts and visualization techniques. Next, and as part of the evaluation of the course, students implemented and presented a project which included 12 types of learning activities: brainstorming, designing visualization, gathering data, parsing, filtering and mining data, getting started with the visualization library D3.js, implementing the visualization, implementing interaction in the visualization, reading resources, reading research papers, preparing questions, and preparing research presentations.

As participants were students registered for an information visualization course, they all had a relevant level of knowledge of principles and theories involved in the creation of visualizations. Thus, their feedback was highly relevant during the stage of early definition and development of the visualizations.

In total, 42 students were registered for the course. Out of these 42 students, ten students participated in the first user study.

5.2 Data collection in user study 1

This pilot study mainly served to identify usability issues of the radial and timeline visualizations and to collect feedback and input from students for additional visualizations. In this study, we first conducted ten think-aloud sessions (Lewis, 1982), with one student at a time. Each session was organized in three phases: filling out a survey to capture emotion-related data about their work during the project, conducting tasks with the two visualizations, such as identifying the most frequent emotion with the visualizations, and filling out an evaluation survey about the visualizations.

The survey that intended to capture data about students' activities during the project used explicit questions about the students' affective state for each type of activity conducted in the project. For each type of learning activity, students had to indicate how frequently they have experienced the five affective dimensions known to occur in learning scenarios: motivation, happiness, boredom, confusion, and frustration (DMello *et al.*, 2007). Students were also asked to indicate the amount of time they dedicated to the project during each week. Afterwards, the two visualizations were presented, i.e. the radial visualization showing the frequency of emotions for each type of activity and the timeline visualization. Both visualizations used the data collected in the previous phase.

After completing the tasks, such as identifying the most frequent emotion and comparing this value to the class average, students filled out an evaluation survey, including questions about the usability (usefulness and insight) of the visualizations. The usability was subsequently measured with the SUS method. Students also rated the two visualizations in the range of "not useful at all" to "very useful."

Students were also asked which other information could be of interest to represent through visualizations. They were asked to rate the utility of five types of information on a five-point Likert scale: types of used resources (e.g. forums, blogs or files), detailed information about one student, comparing actions between two students, detailed statistics of most used resources and information about content creation by students.

5.3 Data analysis results in user study 1

5.3.1 Perceived usefulness. The usability results obtained from the evaluation of the SUS questions resulted in 72.5 points on average, which can be assessed as a positive belief

(Bangor *et al.*, 2008). The timeline was perceived as the most useful by the students (average score above 3.5 on a scale from 1 to 5). The radial visualization was perceived as useful (score above 3 on a scale from 1 to 5).

Students indicated that they are interested in detailed information of one student, comparison of students and information about content creation by students. The information related to the types of used resources and the most used resources were the least prioritized. In Figure 5, we present a set of box plots illustrating the priorities by students given to each option.

5.3.2 Post-study interview results in study 1. The analysis of students' answers to the interview questions provided useful insights for improvement needs for the visualizations. The results revealed that students experienced difficulties in interpreting certain aspects of visualizations. For instance, some students found it difficult to identify the values on the radial bars that were used to visualize affective states per activity. This difficulty was found to be due to user interface related issues such as having adjacent bars with similar colors or a low-level contrast, making them not easily distinguishable in the chart. Furthermore, some students were not able to interpret the meaning of several visual components. For example, there were students who were not able to detect that the class average was represented by solid lines on the timeline visualization. In general, students expressed that they "liked the timeline and the comparison with the class average" and prefer its use in the future. The radial visualization was difficult to understand by some students ("it is hard to see the information of all students," "the red color (of bars representing frustration) is too distracting" or "it is confusing that bars do not start from zero"), which suggested that the interpretability of the visualization needed to be further improved.

Some of the post-study responses provided creative input in the form of suggestions for further information needs of students. For instance, there was a suggestion to include the equivalent of "return over investment," where the dedicated time would represent the investment and the obtained mark would represent the return. Another suggestion was the inclusion of task types to which the time was dedicated, such as lectures, homework, studying, and group work.

6. Extended study with a larger group of students without knowledge in information visualization: user study 2

The second study was conducted with bachelor-level students at the Eindhoven University of Technology in the Netherlands. For the second user study, we improved the two visualizations based on the findings and suggestions of students of the pilot study. To address the difficulties of interpretation, the contrast of colors and the visibility of elements in both visualizations were improved. Interactivity was added to clarify the details of the visualizations: the radial visualization was adjusted to show the value of each bar when the mouse cursor hovered over it. The timeline was adjusted to offer the option for hiding and showing data series, etc. In addition, for this user study, we implemented

Figure 5.
Frequency for answers given to the question "What other data would you like to have visualized or have accessible?"
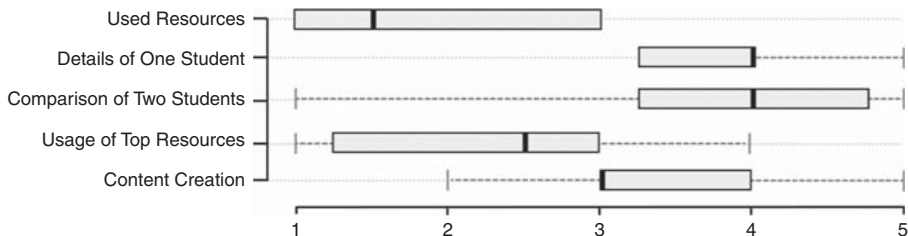
two new visualizations with emphasis on individual and detailed information, as such information was indicated as relevant by students of the first user study. These visualizations are the heatmap and scatterplot visualization described in Section 4. The purpose of this study was to evaluate the usability of the four visualizations, namely, the improved versions of the two visualizations used in user study 1, and the two new visualizations.

6.1 Demographics of participants in user study 2
Overall, 105 first-year bachelor students enrolled in technological programs took part in user study 2. The majority of participants are male (96.6 percent males and 9.4 percent females). Participants are between 18 and 40 years.

6.2 Data collection in user study 2
The study was conducted in the context of the course human-technology interaction. At the beginning of the course, an introduction was provided about all the concepts and processes involved in the design of usable interfaces for technological artifacts. At the end of the semester, students completed a project about the design of a thermostat. The project duration was four weeks, from late April to early June in 2014. In the end, students presented their project to the teaching staff and their peers.

For this project, in collaboration with the instructors, we defined six types of learning activities: brainstorming, interface design, implementation, writing documentation, experiment with users, and writing installation instructions. During the project, students received an e-mail with a link to an online survey each week, as well as a link to a web application that showed the visualizations of their emotion-related data. To maintain the anonymity of information, students were asked to create a personal identifier with which they could access the visualizations. Every week, students completed the following tasks: filling out a survey about their emotions and learning activities during that week, exploring their data in relation to data of other students with the proposed interactive visualizations, and filling out a survey about their perceptions and judgments on the visualizations.

The survey included questions about particular activities. Students were asked to indicate how frequently they had experienced each affective state while performing each of the project activities and the time they dedicated to the project. Students were allowed to report activities for a week different than the current one. After the data were submitted, the student could use a web application to access the visualizations.

The data collected in user study 2 were used as follows:

- Radial visualization: values for each affective emotion and each activity were shown explicitly. The average for all students was computed and shown as a solid line.

- Timeline visualization: values for the student were plotted according to the week they were provided. An average for the class was also included.

- Heatmap visualization: the intensity value of each cell represents the average value for the corresponding emotion for the given week.

- Scatterplot visualization: for each affective state, the visualization plots the values (scores) for each student along the date and time of the survey submissions.

In this study, the students were asked to answer questions of an evaluation survey to assess the usability of the visualizations on a continuous basis. The usability was evaluated through SUS questions, while the usefulness was evaluated using five-point Likert scales that rank each one of the visualizations from "not useful at all" to "very useful." In addition

to these questions, we also used questions to objectively assess the insight of the visualizations as follows:

- five-point Likert scale to indicate whether the student is much below, below, average, above or much above the class average for each emotion and time dedication;
- indicate the most frequent emotion experienced during the project;
- identify the activity that motivated students (the whole class) the most;
- identify the activity that frustrated the student the most; and
- identify the activity during which the student is most different from the rest.

In addition to weekly evaluations, a think-aloud session took place at the end of the course. The think-aloud session was conducted with batches of two to four groups, involving 6 to 12 participants in each session. At the beginning of each session, the four visualizations were briefly explained. Then the students completed the tasks. Afterwards, we asked students feedback and inquired about their interests. Table I shows the questions asked in this final survey.

Overall, we received 298 submissions from 95 students for the data gathering survey, with 91 percent of the responses from male students and 9 percent of females. Most of the submissions (78.5 percent) belonged to students 20 years old or younger, 17.4 percent between 21 and 25, 1.7 percent between 26 and 30, 0.7 percent between 31 and 35, and 1.7 percent between 36 and 40. The survey for weekly evaluations received 218 responses from 85 students, while only 52 students participated in the final evaluation.

6.3 Data analysis results of user study 2
6.3.1 Perceived usefulness. The average SUS score for the set of visualizations was 60.1. Figure 6 presents a boxplot for each week of the study. The average usability stays constant over the weeks.

Question	Correct (%)	Categories
Survey of the last week (n = 35)		
Identify your most frequent emotion during the project	48	5
Identify the activity that motivated students the most	23	6
Identify activity that frustrated you the most	69	6
Identify activity where you differ the most from peers	31	6
Final survey (n = 47)		
How does your confusion evolve along time?	61	3
Identify the week when students were more frustrated?	34	4
Do you think your time dedication affects your boredom?	53	3

Table I.
Percentage of correct responses to the objective questions in the survey

Figure 6.
SUS scores obtained for all visualizations during the four weeks of the user study
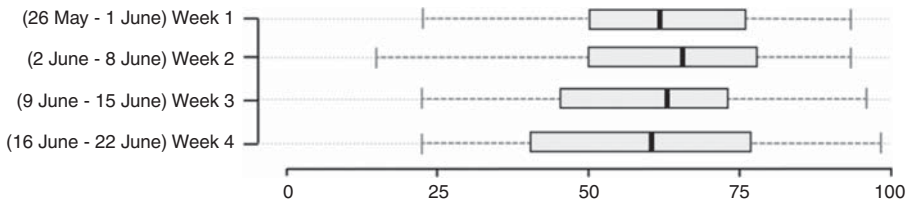
The obtained usability results were found to be lower than in the user study 1. The reason for that can potentially be attributed to the profile of the students. In contrast to the user study 1 participants specialized in visualizations, the students in user study 2 had little or no knowledge about information visualization techniques.

Figure 7 presents the perceived usefulness of each visualization on a five-point Likert scale. The median of the usefulness score was on average 3 for all visualizations. The average was slightly incremented for the timeline and the scatterplot visualization.

6.3.2 Analysis of insight. The differences in student perceptions of their affective states based on the visualizations and the actual values according to data were tested using the Pearson correlation. The analysis, with $n = 34$, resulted in the following findings: frustration ($r = 0.634, p = 0.000$), confusion ($r = 0.620, p = 0.000$), boredom ($r = 0.551, p = 0.000$), happiness ($r = 0.684, p = 0.000$), motivation ($r = 0.829, p = 0.000$), and time dedication ($r = 0.374, p = 0.040$).

For all the relationships, a significant correlation was found ($r > 0.5$), except the time dedication. This suggests that in general students were able to correctly interpret the provided visualizations. However, the results also suggest that there is room for improvement, specifically for the interpretability of visualizations. Ideally, the understanding by all the students would result in higher correlation coefficients closer to 1.

Table I presents the percentage of correct answers based on the students' interpretation of visualizations. These values represent the number of times that the students correctly interpreted the visualizations for different aspects. The number of categories gives an idea of the number of possibilities a student can choose from. These categories are the number of options a student can select to answer a question and are presented in the third column of Table I. For question 1, "identify your most frequent emotion according to visualization," the number of possible answers was limited to 5. We observed that the more options a student has, the more difficult it is for the student to give the right answer. Therefore, the percentages of correct answers should be interpreted taking into account the number of different categories.

Some of the questions included in these surveys also contained a certain level of difficulty that needs to be further discussed. For instance, the affective states with the highest values (motivation and happiness) have a difference of only 0.21. The mean standard deviation for all states is 0.27. As such, in some cases, a student would select an incorrect emotion as his/her "most frequently occurred emotion," however, the value of such emotion was in fact very close to the highest one. The second question of the final survey presents a similar case. Students had to identify the week when they were more frustrated. However, the values for week 2, week 3, and week 4 were similar, with the value of week 4 being just marginally higher than the values of week 2 and week 3. If we had considered all of these three options as valid, 98 percent of the answers would have been correct.

6.3.3 Post-study interview results of user study 2. The analysis of the responses to the interview questions showed that the perceived of the visualization varied among students. Some of them preferred the radial visualization of affective states per learning activity:

> I liked the states per activity the most. After that [I] will go the timeline, followed by the heatmap. Finally the scatterplots.
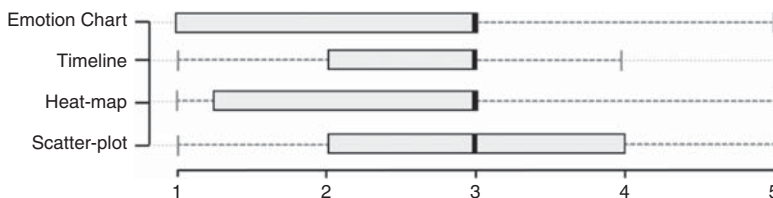

Figure 7.
Usefulness marks from 1 to 5 for each visualization

Other students considered the timeline as the most useful:

> The timeline is the easiest to interpret, since it is in a form I am used to and since it doesn't contain that much data at the same time, which the others do. Especially the heatmap and scatterplot are containing too much detailed and deviating information, which makes it hard to get an overview. The emotion per activity is okay, but also not readable very easily, because some colored areas are very small and it is not always clear which color is represented at what place of the grey line.

Others valued the combination of data and design used in more complex visualizations such as the heatmap:

> It is difficult finding some meaningful values in the scatterplots. However the information in the heatmap is grouped together nicely. Timeline shows a nice overview of how affective states progressed as well.

The teaching staff also provided valuable feedback about potential improvements. The main suggestion was to allow the instructor to indicate an expected amount of time dedication. This would allow students to know whether they were dedicating less (or more) time than what the instructor was planning. The inclusion of expected time dedication would also allow the teaching staff to analyze whether the work load is being set appropriately for the current group of students.

7. Discussion
The results of our study provide useful insights for the usability of different visualizations for students for presenting emotion-related data. However, there are also several limitations that should be articulated. While we were able to assess the usability, measured by perceived usefulness and insight, with a relatively large number of students in user study 2, the limited number of students in user study 1 does not allow to draw strong conclusions from the survey results with this group as the suggestions from this study were mainly used as a basis to improve the visualization design for user study 2.

Second, data collection was performed manually in both user studies. Thus, the accuracy of affective states could be subject to subjective judgments of students. We should, however, mention that, although some studies rely on methods and techniques for capturing and analyzing emotion-related data of students in an automatic way (Leony *et al.*, 2013a, 2015), in this paper we focus on the evaluation of usefulness of representing such data to students. Nevertheless, the acquisition process may also influence perceived usefulness and interpretation of data.

In general, usability results indicate that the visualizations are easy to use for students with knowledge of visualization techniques. A SUS score of 72.5 can be assessed as strongly positive beliefs (Bangor *et al.*, 2008). Although the same results could not be confirmed by user study 2, the results were still found to be acceptable. The average SUS score of 60.1 in this user study still reflects a positive attitude. Since the students who participated in the second study had little or no knowledge of information visualization techniques, the relatively lower scores can be attributed to difficulties with using the visualizations, as can also be inferred from students' answers.

In general, the results suggest that visualization techniques need to be designed with care: the difficulty of interpretation of more complex visualizations, such as the heatmap and radial visualization, may be a barrier for uptake by a general audience with no background in information visualization.

The results on perceived usefulness show that students perceive a simple timeline that represents time dedication and evolution of affective states over time as the most useful visualization. This visualization was rated higher regarding its usefulness than the visualization of affective states per activity in user study 1. In the second user study, with two other visualizations added, this visualization resulted in the most positive scores

on average. In general, the findings suggest that a simpler technique results in a higher perceived usefulness.

Insight was measured by correlations between the actual data indications and student perceptions from visualizations. In addition, we measured how well students were able to interpret the visualizations with the use of objective questions. All the correlations were significantly high, with results higher than 0.5, but still less than 0.7. This suggests that, in the majority of cases, students were able to correctly interpret the provided visualizations.

In summary, the results indicate that visualizations of emotions can support awareness and reflection of student data, but they need to be designed with care to address the needs of students. Simpler techniques, such as timeline visualization, may result in higher positive perceptions than more complex techniques, such as heatmap or radial visualizations. The type of data to include in such visualizations constitutes a further line of research. While students in the user study 1 showed interest in more detailed data about individual students, the representation of such data remains a challenge. Evaluation results of user study 2 indicate that the visualizations that we selected to address their needs (heatmap, scatterplot with three dimensions) are difficult to interpret by users with no background in information visualization. Our future work will focus on exploring visualizations that can represent emotion-related data in a simple and intuitive way to enable use by a general audience.

8. Conclusion

The evaluation presented in this paper showed the potential of dashboards and visualizations to support students awareness of affective information linked to learning activities in an educational scenario. In general, students expressed that they "liked seeing their emotion-related information linked to learning activities and their comparison with their peers."

The results of student evaluations suggest that usability of the proposed visualizations was acceptable, but that there is also room for improvement. In addition, the simpler techniques, such as the timeline visualization, so far offer the highest potential with respect to usability, measured by perceived usefulness and insight. There were differences between students with knowledge and those without knowledge about information visualization. SUS results were higher in user study 1. This suggests that the fact of having knowledge about visualizations might have an influence on the perceived usefulness and insight, and that student training might be necessary in some cases.

Future work includes the improvement and design of new versions of the presented visualizations. Initial modifications will be based on the feedback received during the interviews. Some of these improvements include simplification and adding interaction to ease the interpretation of data. In addition, other visualizations of affective information can be designed to be used by the instructor rather than directly by students.

The work can be ultimately expanded to support integration of this kind of dashboard with emotion detection systems. Applying automatic detection would also provide levels of each emotion in an objective way rather than from a personal perspective, which certainly would improve the validity of the information. Exploring how data that can originate from a multitude of sources and formats can be harvested, curated and fused (Sedrakyan, De Vocht, Alonso, Escalante, Orue-Echevarria and Mannens, 2018) will allow integrating multi-modal data from various wearable sensors and audio/video streams in real-time automated solutions.

In addition, the evaluation presented in this work is limited to perceived usefulness and insight, and does not provide any insight related to potential impact on learning improvements. Thus, expanding the dashboard visualizations with mechanisms to capture a broader scope of learning processes could be another future direction. For instance, process analytics driven approaches (Sedrakyan, 2016; Sedrakyan, De Weerdt and Snoeck, 2016;

Sedrakyan *et al.*, 2014) targeting broader learning-related indicators (Glahn, 2009) will be a relevant future study. Exploring mechanisms for coupling visualizations with textual advice, such as cognitive feedback and behavioral feedforward (Sedrakyan, 2016; Sedrakyan and Snoeck, 2017; Sedrakyan, Järvelä and Kirschner, 2016) as well as a generalizing for different learning goals and tasks (e.g. solo/collaborative learning), is yet another possible direction for future work. Furthermore, not many studies can be found in the domain of feedback automation (Sedrakyan and Snoeck, 2016) thus requiring further research for solid methodologies and frameworks for delivering automated feedback that communicates emotion-related information. Finally, stricter experimental designs with controlling broader evaluation variables and constructs for user acceptance are needed to gain in-depth insights both for future scientific and practical implications.

References

Ali, L., Hatala, M., Gašević, D. and Jovanović, J. (2012), "A qualitative evaluation of evolution of a learning analytics tool", *Computers & Education*, Vol. 58 No. 1, pp. 470-489.

Arnold, K.E. and Pistilli, M.D. (2012), "Course signals at purdue: using learning analytics to increase student success", *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM*, pp. 267-270.

Ashkanasy, N.M. and Dasborough, M.T. (2003), "Emotional awareness and emotional intelligence in leadership teaching", *Journal of Education for Business*, Vol. 79 No. 1, pp. 18-22.

Azcarraga, J., Marcos, N. and Suarez, M.T. (2014), "Modelling EEG signals for the prediction of academic emotions", Workshop on Utilizing EEG Input in Intelligent Tutoring Systems, Honolulu.

Baker, R.S., D'Mello, S.K., Rodrigo, M.T. and 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*, Vol. 68 No. 4, pp. 223-241.

Bangor, A., Kortum, P.T. and Miller, J.T. (2008), "An empirical evaluation of the system usability scale", *Internationall Journal of Human-Computer Interaction*, Vol. 24 No. 6, pp. 574-594.

Barr, J. and Gunawardena, A. (2012), "Classroom salon: a tool for social collaboration", *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education, ACM*, pp. 197-202.

Brooke, J. (1996), "SUS – a quick and dirty usability scale", *Usability Evaluation in Industry*, Vol. 189, p. 194.

Burleson, W. (2006), "Affective learning companions: strategies for empathetic agents with realtime multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance", PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, available at: http://affect.media.mit.edu/pdfs/06.burleson-phd.pdf (accessed February 17, 2018).

Craig, S., Graesser, A., Sullins, J. and Gholson, B. (2004), "Affect and learning: an exploratory look into the role of affect in learning with autotutor", *Journal of Educational Media*, Vol. 29 No. 3, pp. 241-250.

Davis, F.D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly*, Vol. 13 No. 3, pp. 319-340.

D'Mello, S.K., Craig, S.D., Sullins, J. and Graesser, A.C. (2006), "Predicting affective states expressed through an emote-aloud procedure from autotutor's mixed-initiative dialogue", *International Journal of Artificial Intelligence in Education*, Vol. 16 No. 1, pp. 3-28.

DMello, S. (2012), "Monitoring affective trajectories during complex learning", in Seel, N.M. (Ed.), *Encyclopedia of the Sciences of Learning*, Springer, Boston, MA, pp. 2325-2328.

DMello, S., Picard, R. and Graesser, A. (2007), "Towards an affect-sensitive autotutor", *IEEE Intelligent Systems*, Vol. 22 No. 4, pp. 53-61.

Few, S. (2006), *Information Dashboard Design*, O'Reilly.

Friesen, E. and Ekman, P. (1978), *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists, Palo Alto, CA.

GhasemAghaei, R., Arya, A. and Biddle, R. (2016), "A dashboard for affective e-learning: data visualization for monitoring online learner emotions", *Ed-Media: World Conference on Educational Media and Technology*, Vol. 2016, pp. 1536-1543.

Glahn, C. (2009), "Contextual support of social engagement and reflection on the web", SIKS Dissertation Series, Heerlen.

Govaerts, S., Verbert, K., Duval, E. and Pardo, A. (2012), "The student activity meter for awareness and self-reflection", *CHI'12 Extended Abstracts on Human Factors in Computing Systems, ACM, New York, NY*, pp. 869-884.

Greller, W. and Drachsler, H. (2012), "Translating learning into numbers: a generic framework for learning analytics", *Educational Technology and Society*, Vol. 15 No. 42, pp. 42-57.

Janssen, J., Erkens, G., Kanselaar, G. and Jaspers, J. (2007), "Visualization of participation: does it contribute to successful computer-supported collaborative learning?", *Computers & Education*, Vol. 49 No. 4, pp. 1037-1065.

Jaques, P.A. and Vicari, R.M. (2007), "A BDI approach to infer students emotions in an intelligent learning environment", *Computers & Education*, Vol. 49 No. 2, pp. 360-384.

Keim, D.A. (2002), "Information visualization and visual data mining", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 8 No. 1, pp. 1-8.

Kort, B., Reilly, R. and Picard, R.W. (2001), "An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion", *Advanced Learning Technologies, 2001, Proceedings, IEEE International Conference, IEEE*, pp. 43-47.

Leony, D., Muñoz-Merino, P.J., Pardo, A. and Delgado Kloos, C. (2013a), "Modelo basado en HMM para la deteccin de emociones a partir de interacciones durante el aprendizaje de desarrollo de software", *XI Jornadas de Ingeniera Telemática*.

Leony, D., Muñoz-Merino, P.J., Pardo, A. and Delgado Kloos, C. (2013b), "Provision of awareness of learners' emotions through visualizations in a computer interaction-based environment", *Expert Systems with Applications*, Vol. 40 No. 13, pp. 5093-5100.

Leony, D., Parada, G.H.A., Muñoz-Merino, P.J., Pardo, A. and Delgado Kloos, C. (2013), "A generic architecture for emotion-based recommender systems in cloud learning environments", *Journal of Universal Computer Science*, Vol. 19 No. 14, pp. 2075-2092.

Leony, D., Muñoz-Merino, P.J., Pardo, A., Ruiperez-Valiente, J., Arellano Martin-Caro, D. and Delgado Kloos, C. (2015), "Detection and evaluation of emotions in massive open online courses", *Journal of Universal Computer Science*, Vol. 21 No. 5, pp. 638-655.

Lewis, C. (1982), *Using the 'Thinking-Aloud' Method in Cognitive Interface Design*, IBM TJ Watson Research Center.

Maldonado, R.M., Kay, J., Yacef, K. and Schwendimann, B. (2012), "An interactive teachers dashboard for monitoring groups in a multitabletop learning environment", in Cerri, S.A., Clancey, W.J., Papadourakis, G. and Panourgia, K. (Eds), *Intelligent Tutoring Systems*, ITS 2012, Lecture Notes in Computer Science, Vol. 7315, Springer, Berlin and Heidelberg.

Meyer, D.K. and Turner, J.C. (2002), "Discovering emotion in classroom motivation research", *Educational Psychologist*, Vol. 37 No. 2, pp. 107-114.

Mitchell, M.L. and Jolley, J.M. (2012), *Research Design Explained*, Cengage Learning, Wadsworth Publishing, Belmont, CA.

Muñoz-Merino, P.J., Ruiperez Valiente, J.A. and Delgado Kloos, C. (2013), "Inferring higher level learning information from low level data for the khan academy platform", *LAK 2013: Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM, New York, NY*, pp. 112-116.

Muñoz-Merino, P.J., Fernandez Molina, M., Muñoz Organero, M. and Delgado Kloos, C. (2014), "Motivation and emotions in competition systems for education: an empirical study", *IEEE Transactions on Education*, Vol. 57 No. 3, pp. 182-187.

North, C. (2006), "Toward measuring visualization insight", *IEEE Computer Graphics and Applications*, Vol. 26 No. 3, pp. 6-9.

Pardos, Z.A., Baker, R.S.J.D., San Pedro, M.O.C.Z., Gowda, S.M. and Gowda, S.M. (2013), "Affective states and state tests: investigating how affect throughout the school year predicts end of year learning outcomes", *LAK 2013: Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM, New York, NY*, pp. 117-124.

Pekrun, R., Goetz, T., Titz, W. and Perry, R.P. (2002), "Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research", *Educational Psychologist*, Vol. 37 No. 2, pp. 91-105.

Pintrich, P.R. (2003), "A motivational science perspective on the role of student motivation in learning and teaching contexts", *Journal of Educational Psychology*, Vol. 95 No. 4, pp. 667-686.

Pohl, A., Bry, F., Schwarz, J. and Gottstein, M. (2012), "Sensing the classroom: improving awareness and self-awareness of students in backstage", *2012 15th International Conference on Interactive Collaborative Learning (ICL), IEEE*, pp. 1-8.

Ruipérez-Valiente, J.A., Muñoz-Merino, P.J., and Kloos, C.D. (2014), "A demonstration of ALAS-KA: a learning analytics tool for the Khan Academy platform", *European Conference on Technology Enhanced Learning*, Springer, Cham, pp. 518-521.

Ruiz, S., Charleer, S., Urretavizcaya, M., Klerkx, J., Fernandez-Castro, I. and Duval, E. (2016), "Supporting learning by considering emotions: tracking and visualization a case study", *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM*, pp. 254-263.

Santos, J.L., Verbert, K., Govaerts, S. and Duval, E. (2013), "Addressing learner issues with StepUp!: An evaluation", *Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM*, pp. 14-22.

Scheuer, O. and Zinn, C. (2007), "How did the e-learning session go? The student inspector", *Frontiers in Artificial Intelligence and Applications*, Vol. 158, pp. 487-494.

Sedrakyan, G., Snoeck, M. and De Weerdt, J. (2014), "Process mining analysis of conceptual modeling behavior of novices – empirical study using JMermaid modeling and experimental logging environment", *Computers in Human Behavior*, Vol. 41, pp. 486-503.

Sedrakyan, G. (2016), "Process-oriented feedback perspectives based on feedback-enabled simulation and learning process data analytics", PhD thesis, KU Leuven, Leuven, p. 227.

Sedrakyan, G. and Snoeck, M. (2016), "Design and evaluation of feedback automation architecture: enriching model execution with feedback to support testing of semantic conformance between models and requirements", *Proceedings of the 4th International Conference on Model-Driven Engineering and Software Development – Modelsward, Springer*, pp. 14-22.

Sedrakyan, G., De Weerdt, J. and Snoeck, M. (2016), "Process-mining enabled feedback: tell me what I did wrong vs. tell me how to do it right", *Computers in Human Behavior*, Vol. 57, pp. 352-376.

Sedrakyan, G., Järvelä, S. and Kirschner, P. (2016), "Conceptual framework for feedback automation and personalization for designing learning analytics dashboards", *Conference EARLI SIG 27, Online Measures of Learning Processes, Oulu, November 29-December 1*.

Sedrakyan, G. and Snoeck, M. (2017), "Cognitive feedback and behavioral feedforward perspectives for modeling and validation in a learning context", in Hammoudi, S., Pires, L.F., Selic, B. and Desfray, P. (Eds), *Model-Driven Engineering and Software Engineering*, Communications in Computer and Information Science, Vol. 692, Springer International Publishing, Cham, pp. 70-92.

Sedrakyan, G., Leony, D., Munoz-Merino, P., Delgado Kloos, C. and Verbert, K. (2017), "Evaluating student-facing learning dashboards of affective states", *Proceedings of the Twelfth European Conference on Technology Enhanced Learning, Springer*, pp. 1-14.

Sedrakyan, G., De Vocht L., Alonso, J., Escalante, M., Orue-Echevarria, L. and Mannens, E. (2018),
"Data Harvesting, Curation and Fusion Model to Support Public Service Recommendations for
e-Governments", *Proceedings of the 6th International Conference on Model-Driven Engineering
and Software Development, Funchal, Madeira, 22-24 January*.

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S. and Kirschner, P. (2018), "Linking learning
behavior analytics and learning science concepts: designing a learning analytics dashboard for
feedback to support learning regulation", submitted.

Son, L.H. (2012), "Supporting reflection and classroom orchestration with tangible tabletops",
PhD thesis, Ecole Polytechnique Federale de Lausanne, EPFL, Lausanne.

Sylwester, R. (1994), "How emotions affect learning", *Educational Leadership*, Vol. 52 No. 2, pp. 60-65.

Trigwell, K., Ellis, R.A. and Han, F. (2012), "Relations between students' approaches to learning,
experienced emotions and outcomes of learning", *Studies in Higher Education*, Vol. 37 No. 7,
pp. 811-824.

Värlander, S. (2008), "The role of students emotions in formal feedback situations", *Teaching in Higher
Education*, Vol. 13 No. 2, pp. 145-156.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S. and Santos, J.L. (2013), "Learning analytics dashboard
applications", *American Behavioral Scientist*, Vol. 57 No. 10, pp. 1500-1509.

Verbert, K., Govaerts, S., Duval, E., Santos, J., Van Assche, F., Parra, G. and Klerkx, J. (2014),
"Learning dashboards: an overview and future research opportunities", *Personal and Ubiquitous
Computing*, Vol. 18 No. 6, pp. 1499-1514.

Von Alan, R.H., March, S.T., Park, J. and Ram, S. (2004), "Design science in information systems
research", *MIS Quarterly*, Vol. 28 No. 1, pp. 75-105.

Yigitbasioglu, O.M. and Velcu, O. (2012), "A review of dashboards in performance management:
implications for design and research", *International Journal of Accounting Information Systems*,
Vol. 13 No. 1, pp. 41-59.

Corresponding author
Gayane Sedrakyan can be contacted at: gsedrakyan@gmail.com