Affecting Off-Task Behaviour: How Affect-aware Feedback Can Improve Student Learning

Beate Grawemeyer
London Knowledge Lab
Birkbeck, University of London
London, UK
beate@dcs.bbk.ac.uk

Manolis Mavrikis
London Knowledge Lab
UCL Institute of Education
University College London, UK
m.mavrikis@ioe.ac.uk

Wayne Holmes
London Knowledge Lab
UCL Institute of Education
University College London, UK
w.holmes@ioe.ac.uk

Sergio Gutierrez-Santos
London Knowledge Lab
Birkbeck, University of London
London, UK
sergut@dcs.bbk.ac.uk

Michael Wiedmann
Institute of Educational Research
Ruhr-Universität Bochum, DE
michael.wiedmann@rub.de

Nikol Rummel
Institute of Educational Research
Ruhr-Universität Bochum, DE
nikol.rummel@rub.de

ABSTRACT

This paper describes the development and evaluation of an affect-aware intelligent support component that is part of a learning environment known as iTalk2Learn. The intelligent support component is able to tailor feedback according to a student's affective state, which is deduced both from speech and interaction. The affect prediction is used to determine which type of feedback is provided and how that feedback is presented (interruptive or non-interruptive). The system includes two Bayesian networks that were trained with data gathered in a series of ecologically-valid Wizard-of-Oz studies, where the effect of the type of feedback and the presentation of feedback on students' affective states was investigated. This paper reports results from an experiment that compared a version that provided affect-aware feedback (affect condition) with one that provided feedback based on performance only (non-affect condition). Results show that students who were in the affect condition were less bored and less off-task, with the latter being statically significant. Importantly, students in both conditions made learning gains that were statistically significant, while students in the affect condition had higher learning gains than those in the non-affect condition, although this result was not statistically significant in this study's sample. Taken all together, the results point to the potential and positive impact of affect-aware intelligent support.

Categories and Subject Descriptors

J.1 [**Administrative Data Processing**]: Education; K.3.1 [**Computer Uses in Education**]: Computer-assisted instruction

LAK '16, April 25-29, 2016, Edinburgh, United Kingdom
© 2016 ACM. ISBN 978-1-4503-4190-5/16/04...$15.00
DOI: http://dx.doi.org/10.1145/2883851.2883936

Keywords

Affect, Feedback, Exploratory Learning Environments

1. INTRODUCTION

The aim of our research is to provide intelligent support to students, taking into account their affective states in order to enhance their learning experience and performance.

It is well understood that affect interacts with and influences the learning process [19, 10, 3]. While positive affective states (such as surprise, satisfaction or curiosity) contribute towards learning, negative ones (including frustration or disillusionment at realising misconceptions) can undermine learning. Any learning experience is typically full of transitions between positive and negative affective states. For example, while a student may be interested in a particular learning task, any misconceptions that they have might lead to frustration or disillusionment as they are forced to reconsider their existing understanding. However, if this negative affective state is reconciled, the student might become deeply engaged with the task once again. D'Mello et al., for example, elaborate on how confusion, which superficially might be considered a negative affective state, is likely to promote learning under appropriate conditions [10].

It is important therefore, to deepen our understanding of the role of affective states for learning, and to be able to move students out of states that inhibit learning. Pekrun [24] discusses achievement emotions or affective states that arise in a learning situation. Achievement emotions are states that are linked to learning, instruction, and achievement. We focus on a subset of affective states identified by Pekrun: flow/enjoyment, surprise, frustration, and boredom. We also add confusion, which has been identified elsewhere as an important affective state during learning [25], for tutor support and for learning in general [10].

As described in Woolf et al. [32], students can become overwhelmed (very confused or frustrated) during learning, which may increase cognitive load [30]. However, appropriate feedback might help to overcome such problems. Carenini et al. [4] describe how effective support or feedback needs to answer three main questions: (i) when the support should be provided during learning; (ii) what the support should contain; and (iii) how it should be presented.

In this paper we focus on the latter two questions: *what* and *how* support or feedback should be provided based on the student's affective state.

We report on the development of intelligent feedback that is able to tailor the type of feedback as well as the presentation of the feedback in order to enhance a student's learning experience. It includes a Bayesian network for determine the most effective feedback type, as well as a Bayesian network for detecting the most effective presentation of the feedback. Both networks are trained with data from Wizard-of-Oz studies where the impact of a student's affective state on the effectiveness of the feedback type and on the presentation of feedback was investigated (c.f. [15, 13]).

In those studies, we learned that a student's affective state can be enhanced when the feedback type is matched to the affective state of the student. For example, when students were confused, affect boosts and specific instructive feedback were more effective in helping students [15]. Additionally, adapting the presentation of the feedback according to the student's affective state is also important, especially when they are confused or frustrated. For these particular affective states, high-interruptive feedback is more effective, especially as the cost of not viewing the feedback is likely to be a negative affective state [13]. However, when students are in flow, low-interruptive feedback is preferred [21].

In the next section an overview of related literature is provided. Section 3 describes the development of the affect-aware intelligent support. Section 4 outlines the evaluation of the support. Results of the evaluation are reported in section 5. A detailed discussion that highlights the importance of including affect in learner modelling is provided in section 6; while section 7 concludes the paper.

2. RELATED WORK

Most of the related work in the field focuses on detecting affect in different input stimuli, ranging from spoken dialogue (e.g. [31]) to keyboard and mouse interactions [27], including a combination of different stimuli such as conversational cues, body language and facial expression [8].

Only a limited amount of research has been undertaken to investigate how a student's affect or motivation can been taken into account to provide learning material or motivational feedback. One early example is del Soldato & du Boulay [7] who use a student's motivation to decide whether to provide the next task or to provide hints. Another example is Santos et al. [28] who show that affect as well as motivation and self-efficacy impact the effectiveness of motivational feedback and recommendations. Additionally, Woolf et al. [32] developed an affective pedagogical agent which is able to mirror a student's affective state, or acknowledge a student's affective state if it is negative. Another example is Conati & MacLaren [6], who developed a pedagogical agent to provide support according to the affective state of the student and their personal goal. Also, Shen et al. [29] recommend learning material to the student based on their affective state. D'Mello et al. [9] developed a system that is able to respond to students via a conversation that takes into account the affective state of the student.

A limited number of researchers have looked at how the presentation of information or feedback could be adapted according to certain user characteristics. For example, in the area of information visualisation, Carenini et al. [4] describe a study that looks at tailoring visual prompts, based on task complexity, user characteristics (such as perceptual speed, visual working memory, and verbal working memory) and delivery times. Also, Grawemeyer & Cox [12] describe a system that is able to recommend a particular representation (bar chart, plot chart, pie chart, sector graph, eulers diagram, or table) based on the user's expertise with representations, their preferences for particular representations, the task, the information to be presented and the representation's semantics. In the slightly different, but related, use case of exploratory visual data analysis, researchers have looked at inferring the user's intended task and recommending alternative visualisations that may help in their analysis [11]. Further, Ahn & Brusilovsky [1] describe a system which adapts the visualisation of search results dynamically, based on a user's emerging interests.

In this paper we describe how we add to this literature by developing and evaluating intelligent support within a learning environment that is able to adapt the type of feedback as well as the presentation (high- or low-interruptive) of feedback according to a student's affective state in order to enhance their learning experience and learning gains.

3. THE ITALK2LEARN PLATFORM

iTalk2Learn is a learning platform for children aged 8-12 years old who are learning fractions. It combines structured practice with more open-ended activities in an exploratory learning environment called Fractions Lab [22]. The overall aim is to foster robust learning through providing exploratory exercises that help develop conceptual knowledge interleaved with structured practice activities that foster procedural knowledge. In this paper, we focus on the support provided while students are undertaking these learning activities (i.e. not the sequencing of those activities).

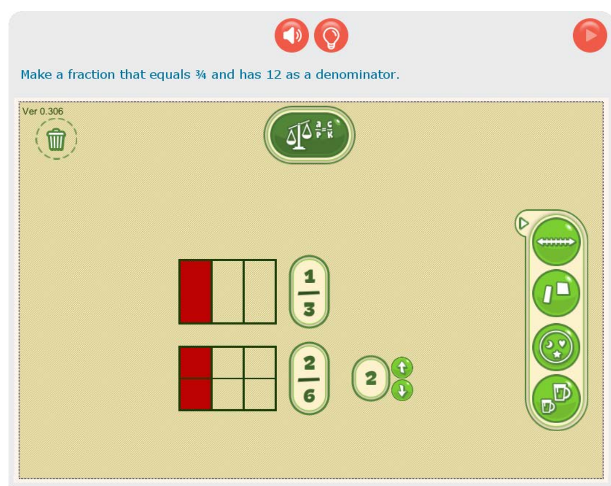

Figure 1: Exploratory learning environment (Fractions Lab).

Figure 1 shows the Fractions Lab interface. The learning task is displayed at the top of the screen. Students are asked to solve the task by choosing fraction representations (from the right-hand side menu) which they manipulate in order to construct an answer to the given task. Adaptive support is provided to the students based on their screen interactions and their speech. The platform is designed to detect and analyse children's speech in near real time (c.f. [17]).

Figure 2 shows the components of the adaptive support. Drawing on our previous work [16] the support consists of three main layers: the analysis or evidence detection layer, the reasoning layer, and the feedback generation layer.

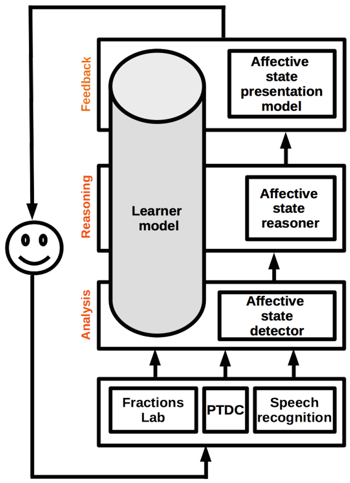

Figure 2: Components of the adaptive support.

The input to the analysis layer includes the interaction with Fractions Lab (e.g. what fraction was generated or changed), the result of the perceived task difficulty classifier (PTDC), which is based on the student's speech (e.g. over-, appropriately or under-challenged), and the data from the speech recognition software (spoken words).

In the analysis layer, the student's interactions with the platform are identified. This layer includes the affective state detector, where a student's affective state is detected from their speech (results from PTDC and speech recognition software) and their interaction with Fractions Lab.

Based on the evidence detection component, the reasoning layer decides if and what feedback should be provided. This layer contains a student model and an affective state reasoner. The student model includes the affective state of the student and information about actions the student has performed, such as representation selected, as well as information about what feedback has been provided to the student and if that feedback was followed. The affective state reasoner uses the information from the student model to decide what type of feedback should be provided.

The feedback generation layer receives the output from the reasoning layer. It includes an affective state presentation model, which is used to decide how the feedback should be presented to the student. The feedback can either be presented in a low-interruptive way (a light bulb glows, indicating that feedback is available, which can then be accessed or ignored by the student) or in a high-interruptive way (the students are interrupted by a pop-up window which has to be dismissed before the student can continue).

Detailed information about the different components are provided below.

3.1 Learner model

The learner model spans all three main components and can be seen as the heart of the intelligent support. It includes the following information about the current student:

- The affective state of the student (that has been calculated by the affective state detector, based on the student's interaction and speech).
- Reasoning stage of the student (beginning, middle or end of the task).
- Interactions with the learning environment (e.g. representation selected or fraction changed).
- The type of feedback that has been provided to the student in the past.
- If the feedback was viewed by the student or not.
- If the feedback provided was followed or not.

This information is used by the different components to determine *what* type of support should be provided to the student, and *how* that support should be provided.

3.2 Analysis layer (affect detection)

The detection of the student's affective state is based on their interaction with the exploratory learning environment (Fractions Lab) as well as on their speech as follows:

- The student's interaction with the platform is used to provide a probability about their affective state. The data used for this calculation is based on whether or not they viewed the feedback and whether or not they followed the advice provided by the feedback. It is a Bayesian network that is trained with data from formative evaluations in Wizard-of-Oz studies [20].
- A perceived task difficulty classifier (PTDC) extracts prosodic features from the student's speech, in order to determine if the student is over-, appropriately or under-challenged [17]. Speech and pause histograms are used by the perceived task difficulty recognition [18].
- The speech recognition software [26] detects whether students are speaking or not and produces an array of spoken words. This array is used to detect certain keywords that are associated with a particular affective state. We apply a naive Bayes classifier for classifying the affective state from those words [14].

The affective state detector determines the overall affective state of the student, based on weightings of the different input components. For example, a higher weight is given to detecting keywords from the speech that are associated with a particular affective state, followed by the interaction data.

3.3 Reasoning layer (affective state reasoner)

The affective state reasoner uses the information from the student model to decide *what* type of feedback should be provided. We explore different types of feedback that are known from the literature to support students' in their learning and that fit our context. The following different feedback types were provided:

- **AFFECT BOOSTS - affect boosts**. As described in [32], affect boosts can help to enhance student's motivation in solving a particular learning task. These included prompts that acknowledged, for example, that a task is difficult or that the student may be confused and encourages them to keep trying.

- **AFFIRMATION PROMPTS - task completion prompts**. This feedback is provided when students have completed the task successfully, in order to give them an indication that they have finished the task and should start working on the next task.

- **INSTRUCTIVE FEEDBACK - instructive task-dependent feedback**. This feedback provided detailed instructions, what subtask or action to perform in order to solve the task.

- **OTHER PROBLEM SOLVING FEEDBACK - task-dependent feedback**. This support was centred on helping students to solve a particular problem that they are facing during their interaction by providing either questions to challenge their thinking or specific hints designed to help them identify the next step themselves.

- **REFLECTIVE PROMPTS - reflecting on task performance and learning**. Self-explanation can be viewed as a tool to help a student address misunderstandings [5] and as a 'window' into a student's thinking.

- **TALK ALOUD PROMPTS - talking aloud**. With respect to learning in particular, the hypothesis that automatic speech recognition (ASR) can facilitate learning is based on educational research that has shown benefits of verbalization for learning (e.g., [2]).

- **TASK SEQUENCE PROMPTS - completing the task**. This feedback is centred on providing support when students try to go to the next task when they have not completed the current task.

Table 1 shows examples of the different feedback types. Based on the information from the learner model, the affective state reasoner decides what type of feedback should be presented to the student. The affective state reasoner is a Bayesian network based on data gathered in Wizard-of-Oz studies [15] that investigated the impact of the different feedback types on the affective state of students. In those studies, students were given a series of fractions tasks and were provided with feedback, of the types described above, by the researchers (the 'wizards') as if it was being provided by the system. The decision about what type of feedback to provide was based on a script. For more information, the reader is referred to [20].

During those studies, the student's affective states were annotated by using the Baker-Rodrigo Ocumpaugh Method Protocol (BROMP) and the HART mobile app that facilitates coding in the classroom [23]. Two researchers also independently annotated the affective states after the Wizard-of-Oz studies using screen and voice recordings. This was then compared against the field annotations. Kappa between the consolidated annotation and the HART data was .71, $p<.05$.

Figure 3 shows the Bayesian network of the affective state reasoner. We trained the network with the data from the Wizard-of-Oz studies (265 data points). For the trained network we employed a 10-fold cross-validation that showed promising results (accuracy = 79.25%; Kappa = 0.50; recall true = 0.62; recall false = 0.87) and that encouraged us to proceed to the implementation of the system proper.

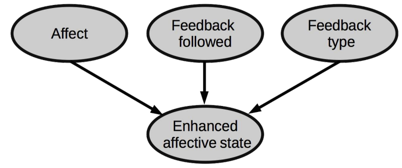

Figure 3: Bayesian network of the affective state reasoner.

The affective state reasoner receives the affective state of the student (based on speech and interaction) as well as information about whether the previous feedback was followed. For each *feedback type*, the enhanced affective state is predicted. This is used to determine which *feedback type* will be most effective at enhancing the affective state at any given time.

3.4 Feedback layer (affective state presentation model)

The aim of the affective state presentation model is to present the feedback in a way that enhances the student's affective state. In our learning environment, the feedback can be presented in a low-interruptive way by highlighting a light bulb, which indicates feedback available (see Figure 4). Additionally, the feedback can be presented in a high-interruptive way by providing a pop-up window that includes the relevant feedback (See Figure 5).

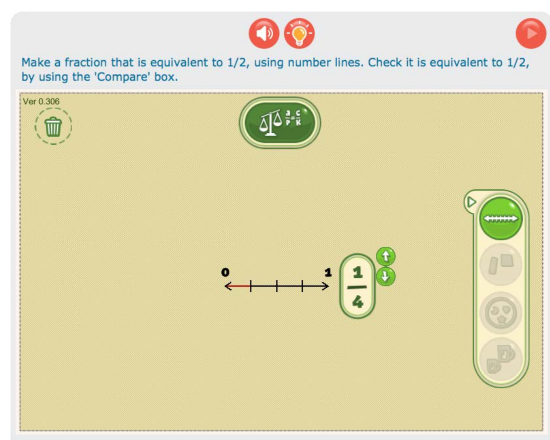

Figure 4: Low-interruptive feedback: the light bulb graphic (at the top of the screen) glows, indicating that feedback is available.

We conducted a second Wizard-of-Oz study that investigated if there was a difference in a student's affective state when the feedback was either presented in a low-interruptive way through the light bulb, or in the high-interruptive way through the pop-up window [13]. Similar to the Wizard-of-Oz studies described earlier, different types of feedback were provided to the students. However, in this study this feedback was either presented in a low-interruptive or in a high-interruptive way. We annotated the students' affective states after the studies based on video and speech data (Kappa = .52, $p<.001$).

The data from the study was used to train a Bayesian network that is able to predict an enhanced affective state

Feedback type	Example
AFFECT BOOSTS	Well done. You're working really hard!
AFFIRMATION PROMPTS	The way that you worked that out was excellent. Now go to the next task.
INSTRUCTIVE FEEDBACK	Use the comparison box to compare your fractions.
OTHER PROBLEM SOLVING FEEDBACK	Good. What do you need to do now, to complete the fraction?
REFLECTIVE PROMPTS	What do you notice about the two fractions?
TALK ALOUD PROMPTS	Please explain what are you doing.
TASK SEQUENCE PROMPTS	Are you sure that you have answered the task fully? Please read the task again.

Table 1: Examples of feedback types

Figure 5: High-interruptive feedback: pop-up window that includes a feedback message.

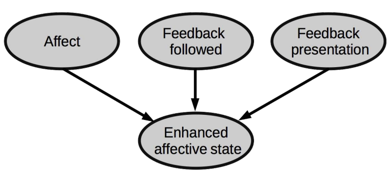

Figure 6: Bayesian network of the affective state presentation model.

by adapting the *presentation* of the feedback. Figure 6 shows the Bayesian network of the affective state presentation model.

The dataset contained 266 cases of affective states that occurred before and after feedback was presented (either high- or non-interruptive) as well as student interaction data (whether or not previous feedback had been followed). Using this data set and employing a 10-fold cross-validation gives encouraging results (accuracy = 82.38%; Kappa = 0.53; recall true= 0.65; false= 0.87).

The Bayesian network of the affective state presentation model is similar to the Bayesian network of the affective state reasoner, but differs in adapting the *presentation* of the feedback rather than the feedback type.

The affective state presentation model receives the affective state of the student as well as information about whether previous feedback was followed. Based on this, the most effective *presentation* of the feedback is detected - one that aims to enhance the affective state of the student.

4. EVALUATION

As mentioned, we were interested in investigating the potential of the learner model and, as an extension, the adaptive support to student learning. The iTalk2learn project ran a series of formative and summative evaluation studies that considered a range of questions. Of relevance here, was a study looking particularly into whether feedback tailored to the affective state of the students enhanced their learning experience and performance. We were particularly interested in the following sub-questions:

- Can speech and interaction be used effectively as inputs to *detect* a student's affective states?
- Can a student's *learning experience* be enhanced when the feedback is tailored to their affective state?
- Can a student's *behaviour* towards the task be enhanced when the feedback is tailored to their affective state?
- Do students have higher *learning gains* when feedback is adapted to their affective state?

To address these questions, we evaluated the system and intelligent support by comparing one version that included the affect-aware support with a version where the affect-aware support was switched off.

4.1 Participants

77 participants took part in the evaluation. They were all primary school students, aged between 8 and 10 years old. Parental consent, for their involvement in the study, was obtained for all students.

4.2 Procedure

The participating students were roughly stratified, according to previous teacher assessments of the children's mathematical ability, and then randomly allocated to two subgroups (approximately equal in size, with each group having approximately the same number of high, middle and low achieving students). The first group (N = 41) was assigned to the affect condition: the students were given access to the full iTalk2Learn system, which uses the student's affective state and their performance to determine the appropriate feedback and its presentation. The second group of students (N = 36) was assigned to the non-affect condition: they were given access to a version of the iTalk2Learn system in which feedback is based on the student's performance only.

Two sessions, one for each condition, were undertaken in each school. At the beginning of each session, students completed an online questionnaire that assessed their knowledge of fractions (the pre-test). This was followed by 40 minutes during which the students engaged with fractions tasks in a version of the iTalk2Learn system that, according to the experimental condition, included either the affect-aware or the non-affect-aware support.

Based on results from our Wizard-of-Oz studies [20], two sets of support were identified that differ in their impact for enhancing a student's learning experience by adapting feedback based on their affective states:

- **No impact for adaptation based on affect:**
  - TALK ALOUD PROMPTS were based on interaction only and were provided in the affect condition only, when students did not say anything for 30 seconds.
  - TASK SEQUENCE PROMPTS were based on interaction only and were provided when students try to go to the next task when they have not completed the current task.
  - AFFIRMATION PROMPTS were based on performance and were provided when students successfully completed the task.

  All of these feedback types were provided in both conditions in a high-interruptive way (pop-up window).

- **High impact for adaptation based on affect:**
  - AFFECT BOOSTS were based on a student's affective state and were provided in the affect condition only.
  - INSTRUCTIVE FEEDBACK, OTHER PROBLEM SOLVING FEEDBACK and REFLECTIVE PROMPTS were tailored based on affect within the affect condition, and based on performance within the non-affect condition.

  In the affect condition the presentation of the feedback (high-interruptive (pop-up) or low-interruptive (light bulb)) was based on the student's affective state. In contrast, in the non-affect condition these feedback types were presented in a low-interruptive way (light bulb). An exception to this was the REFLECTIVE prompts, which were provided in a high-interruptive way at the end of the task in the non-affect group.

While the students engaged with the system, the affective states of a subset of the students' (affect condition: N = 26; non-affect condition: N = 22) were monitored and noted using the Baker-Rodrigo Ocumpaugh Monitoring Protocol (BROMP) and the Human Affect Recording Tool (HART) Android mobile app. BROMP gives strict guidelines on how the affective states of students are detected (by e.g. body posture, facial expression and engagement with the learning environment). The HART mobile app was then used to note the affective states detected with the BROMP protocol.

After the 40 minutes, the students completed a second online questionnaire that again assessed their knowledge of fractions (a post-test similar to the pre-test).

5. RESULTS

5.1 Affect detection

In the affect condition, the students' affective states were detected automatically by the system, as it analysed their speech and their interaction as described earlier. Additionally, the affective states were annotated by a researcher using the HART mobile app with the BROMP method, also as described earlier. The affective states that were detected automatically include *flow, confusion, frustration, boredom*, and *surprise*. With the BROMP method the affective states that the researchers detected included the automatically detected ones and two additional affective state: *delight* and *eureka*.

Both of those data sources include time stamps, identifying when the particular affective state occurred. The affective state from the automatic detection and the HART annotations were matched according to their time stamp (with a 30 seconds window).

There was a moderate agreement between the automatic detection and the HART annotations, Kappa = .53, $p<.001$ (74.07% agreement). The difference is partly due to the two affective states that were detected with the HART tool but that were not included in the automatic detection (i.e. *delight* and *eureka*). Additionally, we knew from our formative phase that *surprise* and *boredom* are difficult to detect automatically and/or rare. Excluding those affective states, a high agreement between the automatic detection and the HART annotations is achieved, Kappa = .62, $p<.001$ (80.00% agreement). However, this result should be read with caution. We ignore the human annotation that the system cannot detect and assume that the annotated states either side of the ignored states are less transient, which together probably suggest a higher agreement than is really the case. Nevertheless, the result is acceptable, especially given that the effect of a misclassification is an intervention that in the best case can help the student and in the worst case can be ignored.

5.2 Adapting feedback message types

In order to investigate differences between the conditions in respect to the adaptations, we outline differences in adapting the types of feedback messages below. In the affect condition, 1971 feedback messages were provided to students. In the non-affect condition students received 2007 messages. Figure 7 shows how these messages were distributed across the different feedback types.

In order to investigate differences between the two conditions (affect and non-affect), a multivariate ANOVA was conducted for the different feedback types. Using Pillai's trace, there was a significant effect of the condition on the number of different types of feedback messages received, V = .929, $F(5,71) = 187.045$, $p < .001$. Separate t-tests on each feedback type revealed significant effects of adapting message type based on affect. There was a difference in how often AFFIRMATION prompts were provided between the affect (M = 2.51, SD = 2.09) and the non-affect (M = 5.33, SD = 2.41) group ($t(75) = -5.50$, $p < .001$, $d = -1.25$). There was also a large difference in how much INSTRUCTIVE

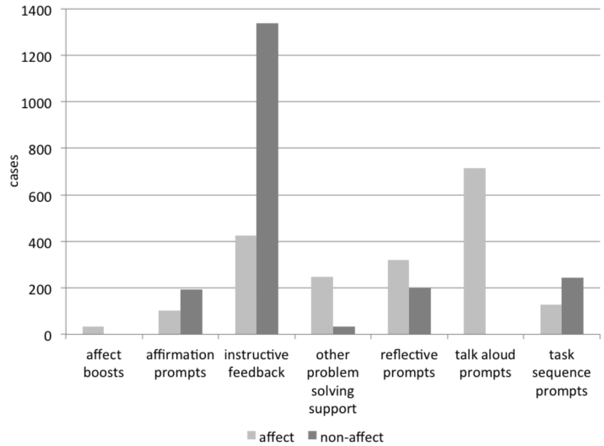

Figure 7: Feedback types provided in the affect and non-affect condition.

feedback was provided between the affect (M = 10.32, SD = 7.04) and the non-affect (M = 37.14, SD = 11.75) group ($t(55.703) = -11.94$, $p < .001$, $d = -2.769$). There were also large differences between conditions for OTHER PROBLEM SOLVING support (affect: M = 6.05 , SD = 2.55 ; non-affect: M = 0.97 , SD = 2.21; $t(74.991) = 9.36$, $p < .001$, $d = 2.129$), REFLECTIVE prompts (affect: M = 7.80, SD = 3.49 ; non-affect: M = 5.53 , SD = 2.21; $t(68.501) = 3.46$, $p < .001$, $d = 0.846$), and TASK SEQUENCE prompts (affect: M = 3.12 , SD = 2.60 ; non-affect: M = 6.78 , SD = 4.22; $t(56.679) = -4.50$, $p < .001$, $d = 1.044$). All of these results were statistically significant.

As described earlier, AFFECT BOOSTS and TALK ALOUD prompts were only provided in the affect condition (affect boosts: M = 0.80, SD = 1.40; talk aloud prompts: M = 17.46, SD = 5.92) and thus could not be compared with the non-affect condition.

5.3 Adapting presentation of feedback

As described earlier, the feedback message was either displayed in a low-interruptive (light bulb) or in a high-interruptive way (pop-up). The way in which the feedback was displayed depended on the impact for adapting feedback based on affective states and, if they were in the affect condition, on the student's affective state.

When feedback was low-interruptive (a glowing light bulb), students could either click on the light bulb and receive the feedback or they could ignore the light bulb and not receive the feedback. In the affect condition, 955 feedback messages were ignored (M = 23.00, SD = 7.54). In the non-affect condition, students ignored 1044 feedback messages (M = 29.00, SD = 11.05). An independent t-test showed that students ignored fewer messages in the affect condition than in the non-affect condition ($t(60.624) = -2.61$, $p < .05$, $d = -0.634$), a result that was statistically significant.

5.4 Affective states and task behaviour

We were particular interested in identifying if a student's affect and their task behaviour can be enhanced through adapting the feedback (type and presentation) according to their affective states.

As described earlier, for a subset of students (affect condition: N = 26; non-affect condition: N = 22) the affective states and task behaviour were monitored by using the Baker-Rodrigo Ocumpaugh Monitoring Protocol (BROMP) and the Human Affect Recording Tool (HART) Android mobile app [23]. For each student, a set of affective states and task behaviour was annotated. Based on these annotations, the percentage of time that a student was in a particular affective state and their task behaviour were calculated. This was used for further analysis as described below.

5.4.1 *Affect*

Figure 8 shows the different types of affective states that were detected in the affect and non-affect condition.

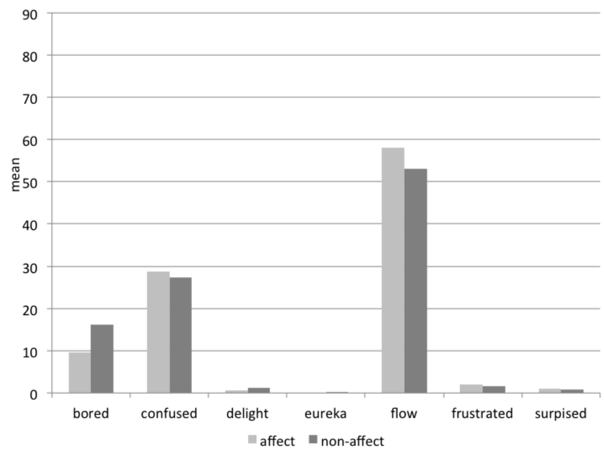

Figure 8: Students' affective states during the main evaluation session in the affect and non-affect condition.

In both conditions students were mainly in *flow* (affect: M = 58.12, SD = 22.23; non-affect: M = 52.98, SD = 17.41; d = 0.257). This was followed by *confusion* (affect: M = 28.77, SD = 23.28; non-affect: M = 27.36, SD = 18.21; d = 0.067) and *boredom* (affect: M = 9.54, SD = 13.33; non-affect: M = 16.08, SD = 7.45, d = -0.606). Only a few were *frustrated* (affect: M = 2.01, SD = 3.15; non-affect: M = 1.54, SD = 2.36; d = 0.169), *surprised* (affect: M = 1.03, SD = 1.83; non-affect: M = 0.74, SD = 2.07; d = 0.148), or *delighted* (affect: M = 0.53, SD = 1.33; non-affect: M = 1.19, SD = 2.50; d = -0.33).

In order to investigate differences between the two conditions (affect and non-affect), a multivariate ANOVA was conducted for the different affective states. Using Pillai's trace, no significant effect of the condition on the affective states could be detected, $V = .188$, $F(6,41) = 1.586$, $p > .05$. However, the medium effect size in *boredom* (d = 0.606) can be seen as an indicator that students were indeed less bored within the affect condition, just not significantly so in this sample. The medium effect can be seen as an indicator that we might find this effect in other samples. Adapting feedback based on affect can decrease boredom.

5.4.2 *Task behaviour*

Figure 9 shows the different types of behaviour that occurred during the evaluation.

In both conditions, students were mainly *on task* (affect: M = 83.58, SD = 13.33; non-affect: M = 82.42, SD = 8.29,

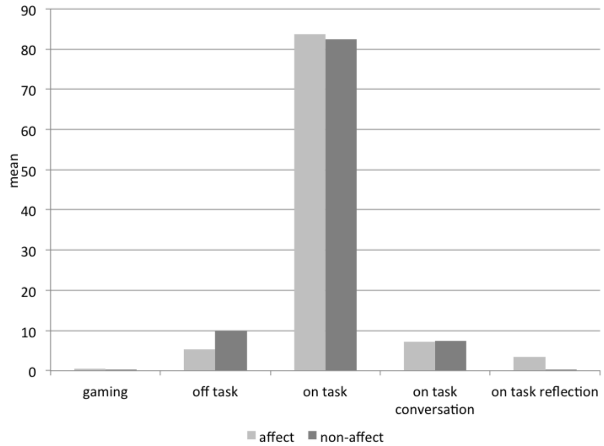

Figure 9: Students' task behaviour during the main evaluation session in the affect and non-affect condition.

d = 0.105). Fewer students did have an *on task conversation* (affect: M = 7.24, SD = 7.86; non-affect: M = 7.36, SD = 6.02, d = -0.017), were *off-task* (affect: M = 5.39, SD = 6.48; non-affect: M = 9.87, SD=6.03, d = -0.716), or *reflecting* on the task (affect: M = 3.38, SD = 9.86; non-affect: M = 0.23, SD = 0.75, d = 0.451). Very few were *gaming* the system (affect: M = 0.41, SD = 1.45; non-affect: M = 0.12, SD = 0.55, d = 0.264).

In order to investigate differences between the two conditions (affect and non-affect), two multivariate ANOVAs were conducted: one for off-task, and one for on-task behaviours. Using Pillai's trace, no significant effect of the condition on on-task behaviours could be detected, V = .125, $F(3,44) = 2.094$, $p > .05$. However, there was a significant effect of condition on *off-task* behaviours, V = .135, $F(2,45) = 3.519$, $p < .05$. Follow-up t-tests revealed a large difference on students' *off-task* behaviour. Students in the affect condition were less *off-task* than students' in the non-affect condition, $t(46) = -2.46$, $p<.05$, d = -.716, a result that was statistically significant.

5.5 Learning gains

Figure 10 shows the students' performance when answering fractions tasks before and after they had used the learning environment in the different conditions.

In the affect condition students increased their knowledge of fractions from M = 2.49 (SD = 1.65) to M = 3.83 (SD = 1.46). In the non-affect condition students increased their knowledge from M = 2.44 (SD = 1.58) to M = 3.33 (SD = 1.71). An ANOVA repeated measures showed an increase of knowledge in both groups ($F(1,75) = 43.94$, $p<.001$, $\eta_p^2 = .369$), a result that was statistically significant.

Although, the difference in learning gains between the groups was not significant ($F(1,75) = 1.81$, $p>.05$), the overall tendency of the affect condition showing higher learning gains warrants further investigation.

6. DISCUSSION

The results of our evaluation will be discussed in respect to our main research questions.

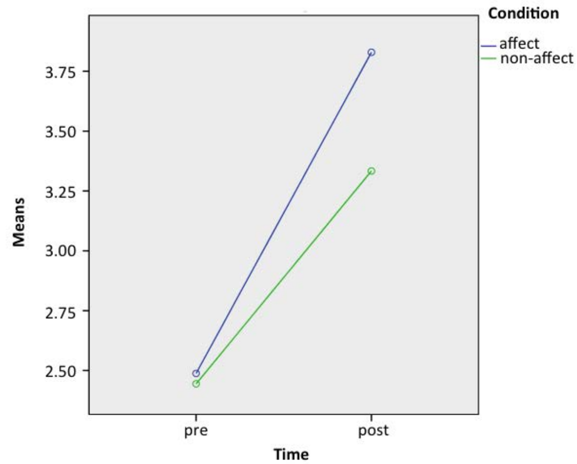

Figure 10: Student's learning gains in the affect and non-affect condition.

6.1 Can speech and interaction be used as input to detect a student's affective states?

In our system, the automatic detection of a student's affective states is based on their speech and their interaction with the learning environment. This automatic detection was compared to the affective states that were annotated with the HART mobile app using the BROMP method (where the affective states are detected by looking at a student's facial expression, body posture and engagement with the learning environment). The comparison revealed a medium agreement when taken into account all affective states (*flow, confusion, frustration, boredom, surprise, delight* and *eureka*).

Out of those affective states *delight* and *eureka* were not included in the automatic affect detection. Additionally, we knew from our formative phase that *surprise* and *boredom* are difficult to detect automatically and are rare. Other researchers (e.g. [3]) also reported on the low frequency of some affective states, such as *surprise*. Excluding these affective states a high agreement between the human annotation and automatic detection was indicative of the quality of the latter.

Nevertheless, our analysis shows that the detection of affective states using information from speech and interaction data is very promising and warrants further research.

6.2 Can a student's learning experience be enhanced when the feedback is tailored to their affective state?

In both conditions, students were mainly in *flow*, followed by *confusion*. This is similar to results from other research, such as [3]. *Flow* and *confusion* contribute towards learning and can be seen as positive affective states.

Students in the affect condition were less *bored* than students in the non-affect condition. While this effect was not significant, given the exploratory nature of this work, we see this as a first indicator that tailoring feedback to affective states can enhance students' learning experience. Future analyses will investigate how tailoring feedback could enhance the students' learning experience. For example: was tailoring of feedback able to transfer students from a negative affective state (such as it boredom) into a positive affective state (such as *flow*)?

The adaptation of the presentation of the feedback might also have been important, as students in the non-affect condition might have ignored feedback presented in the low-interruptive way and therefore might have moved from e.g. *confusion* state to a negative affective state, e.g. *boredom*. This needs further investigation.

6.3 Can a student's behaviour towards the task be enhanced when the feedback is tailored to their affective state?

Students in both conditions were mainly *on task*. This might be due to the nature of the exploratory learning environment, which appeared to engage the students.

However, there was a difference between the groups in *off-task* behaviour. Students in the non-affect group were more *off-task* than students in the affect condition, a result that was statistically significant. Here, the adaptations of the feedback types as well as the presentation of the feedback based on the student's affective state seem to have had an effect on their engagement with the task. Students that are *bored* or *frustrated* might show *off-task* behaviour. It looks as if the adaptations based on a student's affect are able to reduce such negative affective states, which reduce *off-task* behaviour.

Anecdotal evidence from class observations and discussions among and with students suggests that students might have found the task more interesting when feedback was adapted according to their affective states.

6.4 Do students have higher learning gains when feedback is adapted to their affective state?

In both groups, student knowledge of fractions was enhanced. This is a positive result that demonstrates the quality of the tasks and the overall level of support even without affect-aware feedback. However, looking at the result in terms of the impact of the affect detection, the difference between the groups in how much knowledge they gained was not statistically significant but the higher-learning gains in combination with the rest of our results and particularly the significantly higher self-reported interest are encouraging.

As described earlier, the adaptation of the presentation of the feedback might have been important for an increase in learning gains, as students in the non-affect condition might have ignored feedback that was presented in the low-interruptive way.

Also, the range of different types of feedback was spread more evenly in the affect condition than in the non-affect condition. In the non-affect condition mainly INSTRUCTIVE FEEDBACK was provided, based on the students' performance. Meanwhile, in the affect condition the feedback types were tailored according to the students' affect, and much more OTHER PROBLEM SOLVING FEEDBACK and REFLECTIVE PROMPTS were provided. Also interesting to see is that there were fewer TASK SEQUENCE PROMPTS in the affect condition than in the non-affect condition, which indicates that fewer students attempted to skip the task in the affect condition than in the non-affect condition.

This indicates that the affect aware support is able to lead students into positive affective states, such as *flow* where they tend to benefit from reflective or non-instructive more open-ended problem solving feedback.

Because of the different types of feedback, students might have gained different types of knowledge in the different conditions. More INSTRUCTIVE FEEDBACK was provided to students in the non-affect condition, which might have led to an emphasis on procedural knowledge gains. Meanwhile, more OTHER PROBLEM SOLVING FEEDBACK and REFLECTIVE PROMPTS were provided to students in the affect condition, which might have led to conceptual knowledge gains. This again warrants further investigation.

7. CONCLUSION

We have developed an adaptive environment that provides intelligent support according to a student's affective states. It includes two Bayesian networks, which are able to predict an enhancement in a student's affective state by adapting the *type* of feedback and the *presentation* of the feedback according to their affective states.

The intelligent affect-aware support is included in a learning platform, where it can be switched on or off. This feature was used to evaluate the affect-aware support (affect condition), by comparing it to support that was based on performance only (non-affect condition). During our evaluation the students' affective states were annotated while they were using the system in either conditions. The results show that students in the affect condition showed less *off-task* behaviour then students in the non-affect condition, a result that was statistically significant. Additionally, the results indicate that in the affect-aware condition students were less *bored* than students in the non-affect condition. These are important findings as off-task behaviour and boredom can have a negative impact on learning.

The results also underpin the effectiveness of the performance of the training data of the Bayesian networks, as students in the affect condition were more engaged (less *off-task* and less *bored*) than students in the non-affect condition.

Future work includes the refinement of the Bayesian networks with the newly collected data. Additionally, we plan to analyse our data further by looking at a student's affective states and their interactions with the learning environment.

8. ACKNOWLEDGMENTS

This research has been funded by the European Union in the Seventh Framework Programme (FP7/2007-2013) in the iTalk2Learn project (318051).

9. REFERENCES

[1] J. Ahn and P. Brusilovsky. Adaptive visualization for exploratory information retrieval. *Information Processing and Management*, 49:1139–1164, 2013.

[2] M. Askeland. Sound-based strategy training in multiplication. *European Journal of Special Needs Education*, 27(2):201–217, 2012.

[3] R. S. J. d. Baker, S. K. D'Mello, M. T. Rodrigo, and A. C. Graesser. 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. *Int. J. Hum.-Comput. Stud.*, 68(4):223–241, apr 2010.

[4] G. Carenini, C. Conati, E. Hoque, B. Steichen, D. Toker, and J. Enns. Highlighting interventions and

user differences: Informing adaptive information visualization support. In *CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1835–1844, 2014.

[5] M. Chi. Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser, editor, *Advances in instructional psychology*, pages 161–238. Mahwah, NJ: Lawrence Erbaum Associates, 2000.

[6] C. Conati and H. MacLaren. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 2009.

[7] T. del Soldato and B. du Boulay. Implementation of motivational tactics in tutoring systems. *Journal of Artificial Intelligence in Education*, 6(4), 1996.

[8] S. D'Mello and A. Graesser. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-adapted Interaction*, 20:147–187, 2010.

[9] S. D'Mello and A. Graesser. Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, 2(4):1–38, 2013.

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

[11] D. Gotz and Z. Wen. Behaviour driven visualization recommendation. In *Intelligent User Interfaces, IUI '09*, pages 315–324, 2009.

[12] B. Grawemeyer and R. Cox. Graphical data displays and database queries: Helping users select the right display for the task. In *Smart Graphics, 5th International Symposium*, pages 53–64, 2005.

[13] B. Grawemeyer, W. Holmes, S. Gutiérrez-Santos, A. Hansen, K. Loibl, and M. Mavrikis. Light-bulb moment? towards adaptive presentation of feedback based on students' affective state. In *IUI 2015*, 2015.

[14] B. Grawemeyer, M. Mavrikis, A. Hansen, C. Mazziotti, and S. Gutiérrez-Santos. Employing speech to contribute to modelling and adapting to students' affective states. In *EC-TEL 2014*, 2014.

[15] B. Grawemeyer, M. Mavrikis, W. Holmes, A. Hansen, K. Loibl, and S. Gutiérrez-Santos. Affect matters: Exploring the impact of feedback during mathematical tasks in an exploratory environmen. In *AIED 2015*, 2015.

[16] S. Gutiérrez-Santos, M. Mavrikis, and G. Magoulas. A separation of concerns for engineering intelligent support for exploratory learning environments. *Journal of Research and Practice in Information Technology*, 44(3):347–360, 2012.

[17] R. Janning, C. Schatten, and L. Schmidt-Thieme. Feature analysis for affect recognition supporting task sequencing in adaptive intelligent tutoring systems. In *EC-TEL 2014*, 2014.

[18] R. Janning, C. Schatten, and L. Schmidt-Thieme. Perceived task-difficulty recognition from log-file information for the use in adaptive intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, in press.

[19] B. Kort, R. Reilly, and R. Picard. An affective model of the interplay between emotions and learning. In *IEEE International Conference on Advanced Learning Technologies*, number 43-46, 2001.

[20] M. Mavrikis, B. Grawemeyer, A. Hansen, and S. Gutiérrez-Santos. Exploring the potential of speech recognition to support problem solving and reflection - wizards go to school in the elementary maths classroom. In *EC-TEL 2014*, pages 263–276, 2014.

[21] M. Mavrikis, S. Gutiérrez-Santos, E. Geraniou, and R. Noss. Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments. *Personal and Ubiquitous Computing*, 17(8):1605–1620, 2013.

[22] C. Mazziotti, W. Holmes, M. Wiedmann, K. Loibl, N. Rummel, M. Mavrikis, A. Hansen, and B. Grawemeyer. Robust student knowledge: Adapting to individual student needs as they explore the concepts and practice the procedures of fractions. In *Workshop on Intelligent Support in Exploratory and Open-ended Learning Environments Learning Analytics for Project Based and Experiential Learning Scenarios at AIED 2015*, 2015.

[23] J. Ocumpaugh, R. Baker, and M. Rodrigo. Baker-rodrigo observation method protocol (bromp) 1.0. training manual version 1.0. Technical report, New York, NY: EdLab. Manila, Philippines: Ateneo Laboratory for the Learning Sciences., 2012.

[24] R. Pekrun. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *J. Edu. Psych. Rev.*, pages 315–341, 2006.

[25] K. Porayska-Pomsta, M. Mavrikis, and H. Pain. Diagnosing and acting on student affect: the tutor's perspective. *User Modeling and User-Adapted Interaction*, 18(1):125–173, 2008.

[26] Sail-Labs. http://www.sail-labs.com.

[27] S. Salmeron-Majadas, O. Santos, and J. Boticario. Exploring indicators from keyboard and mouse interactions to predict the user affective state. In *EDM 2014*, 2014.

[28] O. Santos, M. Saneiro, S. Salmeron-Majadas, and B. J.G. A methodological approach to elicit affective educational recommendataions. In *IEEE 14th International Conference on Advanced Learning Technologies*, 2014.

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

[30] J. Sweller, J. G. van Merrienboer, and G. W. Paas. Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10:251 – 296+, 1998.

[31] T. Vogt and E. André. Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In *Multimedia and Expo (ICME05)*, pages 474–477, 2005.

[32] B. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. Cooper, and R. Picard. Affect-aware tutors: recognising and responding to student affect. *Int. J. Learning Technology*, 4(3-4):129–164, 2009.