Impact of Interaction Context on the Student Affect-Learning Relationship in Child-Robot Interaction

Huili Chen
MIT Media Lab
hchen25@media.mit.edu

Hae Won Park
MIT Media Lab
haewon@media.mit.edu

Xiajie Zhang
MIT Media Lab
xiajie@media.mit.edu

Cynthia Breazeal
MIT Media Lab
cynthiab@media.mit.edu

ABSTRACT

Prior work in affect-aware educational robots has often relied on a common belief that the relationship between student affect and learning is independent of agent behaviors (child's/robot's) or unidirectional (positive/negative but not both) throughout the entire student-robot interaction. We argue that the student affect-learning relationship should be interpreted in two contexts: (1) social learning paradigm and (2) sub-events within child-robot interaction. In our paper, we examine two different social learning paradigms where children interact with a robot that acts either as a tutor or a tutee. Sub-events within child-robot interaction are defined as task-related events occurring in specific phases of an interaction (e.g., *when the child/robot gets a wrong answer*). We examine sub-events at a macro level (entire interaction) and a micro level (within specific sub-events). In this paper, we provide an in-depth correlation analysis of children's facial affect and vocabulary learning. We found that children's affective displays became more predictive of their vocabulary learning when children interacted with a tutee robot who did not scaffold their learning. Additionally, children's affect displayed during micro-level events was more predictive of their learning than during macro-level events. Last, we found that the affect-learning relationship is not unidirectional, but rather is modulated by context, i.e., several affective states facilitated student learning when displayed in some sub-events but inhibited learning when displayed in others. These findings indicate that both social learning paradigm and sub-events within interaction modulate student affect-learning relationship.

KEYWORDS

affective computing; pedagogical agent; social robotics; human-robot interaction

HRI '20, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-6746-2/20/03…$15.00
https://doi.org/10.1145/3319502.3374822

ACM Reference Format:
Huili Chen, Hae Won Park, Xiajie Zhang, and Cynthia Breazeal. 2020. Impact of Interaction Context on the Student Affect-Learning Relationship in Child-Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3319502.3374822

1 INTRODUCTION

A considerable amount of research has demonstrated the correlation between affect and learning, and indicates that student affect has the potential to foster or hinder learning depending on the interaction context when it is elicited [9, 12, 24, 27]. Hence, it is important to establish a contextualized understanding of the relationship between affect and learning. Guidelines that capture the affect-learning relationship in different contexts would be very useful in the design of affective pedagogical robots. However, prior work in affective pedagogical robots seldom interpret students' affective displays with respect to different social learning paradigms or in different interaction contexts [15, 26, 31, 33]. Instead, a single model or policy is often applied across different contexts. Over-simplification also occurs, i.e., affect observations are only made during certain contexts.

Our paper aims to empirically investigate the correlation between students' learning outcomes with respect to their facial affect in different social learning paradigms (i.e., learning with a robot tutor and with a robot tutee), and varying levels of sub-events within child-robot interaction from macro (entire interaction) to micro (within specific interaction event). We collected a dataset of 40 kindergarten-age children who played a collaborative education game with a social robot companion. The game app was designed to help children learn new vocabulary concepts using a touch-screen tablet. Children played with a robot that either behaved as a more-knowledgable tutor or as a less-knowledgable tutee for two 30-min learning sessions.

This paper makes the following contributions. To our knowledge, it is the first comparative study to investigate how child-robot interaction paradigms (i.e., tutee/tutor robots) and varying levels of sub-events within child-robot interaction impact children's affect-learning relationships. Second, we present a comprehensive evaluation method to extract children's facial affect and analyze student affect-learning relationship. Lastly, we provide analyses and proofs that both robot's social role and sub-events within interaction modulate children's affect-learning relationship.

2 RELATED WORK

2.1 Affect and Learning

Positive affect (e.g., concentration, joy and excitement) is believed to facilitate learning by strengthening motivation [27], increasing persistence [29], improving the use of mental resources [2], and facilitating learning-relevant cognitive processes such as information processing and category sorting tasks [14]. In contrast, negative affect (e.g., frustration, boredom and anger) can be detrimental to learning [9, 27], leading to decreased motivation and effort [23, 28]. However, negative affect states can also facilitate learning under certain situations [12, 24]. For example, confusion can promote learning by increasing the focus of attention on learning material [12]. Mild and acute stress can improve learning and memory [36]. Sometimes negative states are more beneficial than positive ones. For instance, Mills et al. [24] found that people's deep-reasoning comprehension on expository text became better when they felt sad than happy. Taken as a whole, both negative and positive affective states can benefit learning depending on the interaction context in which they are elicited.

2.2 Affect-Aware Educational Robot

Human facial expression is one of the most powerful channels to sense and detect affective due to the face's rich expressiveness and relative ease of automatic data collection and analysis by video [4, 11]. Children's facial expressions have been widely integrated into affect-based interactive learning methods using social robots and virtual avatars [6, 10, 15, 33, 37]. However, analyzing the triadic interplay between the interaction context, elicited student affect, and learning is widely unaddressed in the field.

Prior research has often assumed a fixed, uni-directional relationship between an affective state and learning outcomes across an entire learning interaction – without examining the modulatory effect of interaction context nor incorporating this interdependence into agent behavior models [15, 26, 37]. For example, Woolf et al. [37] implemented a simple rule-based affect extension in their virtual tutoring system to deal with students' affective states. Other works have used model-free reinforcement learning for a robot to learn a single personalized policy to foster learning and engagement [15, 26]. In these systems, the policy's reward function (i.e., a weighted sum of student valence and engagement) remains fixed across the entire child-robot game play – irrespective of learning progress or the child's and robot's behaviors. In other words, the child's positive valence is always treated as a positive reinforcer for the robot's action, and negative valence is always used as a negative reinforcer.

Prior work has also tended to privilege a fixed set of affective states when constructing student models to design robot's reactions to learners' affect [15, 26, 33]. For example, Spaulding et al. [33] incorporated affective data into student skill estimation models using a Bayesian Knowledge Tracing (BKT) algorithm to infer a child's acquisition of reading skills. However, it assumed that smile and engagement states (measured by Affdex) were the most relevant to children's learning. The model incorporated only these two affective measures (over more than 10 other affect and facial expression features) as the observation nodes in their BKT model. An empirical analysis of the possible relationships between multiple affect features and student learning outcomes was not provided in their child-agent context.

Recently, there is a growing appreciation that face-based affect should be recognized and interpreted in the context of interaction [5, 6, 17, 31] . For example, Hoque et al. [17] found that users can smile in response to frustration during a task deliberately designed to elicit the users' frustration. However, widely-used affect recognition tools only recognize one type of smile (assuming positive valence). To remedy this, Hoque et al. [17] integrated contextual information from the interaction (such as instances of incorrect input) into their affect recognition model to achieve a higher accuracy of classifying frustrated verses happy smiles. Despite the growing attention on contextualized interpretation of affect, it seems that affect integration rules are still predominantly determined by theory, experts, and intuition. We argue that data-driven empirical investigations on the complex links between interaction context, student affect and learning are also needed.

2.3 Robot Role and Macro/Micro Events Within Child-Robot Interaction

Social robots have been used to support children's learning in a variety of learning contexts such as storytelling [26], second language learning [20, 30, 32] and STEM [3, 8, 21]. In these contexts, the robot often takes on either a tutor role or a tutee role [1]. Prior work found that a tutee robot elicited greater enjoyment in kids than a tutor robot [19], but a tutee robot might hinder children's learning under some circumstances, which can lead to frustration/confusion [35]. In addition, [7] found that children interacting with a tutor robot showed less face-based affect than children with a tutee robot. The authors, however, did not investigate the interplay between children's affect and learning in each of the tutor and tutee groups. Therefore, this work will examine the impact of robot's role (tutor/tutee) on both children's affective display and affect-learning relationship.

Regarding sub-events within child-robot interaction, prior work often compared and analyzed affect signals over an entire interaction [7, 15, 26]. For example, [7] analyzed children's affect over the entire learning interaction but neglected children's affective display during micro-level events (e.g., robot's actions) within an interaction. In this work, we argue that aggregating children's affect exhibited across a macro event can very likely lose the subtlety of children's affect, especially for those transitory facial expressions/affective states (e.g., dimpler).

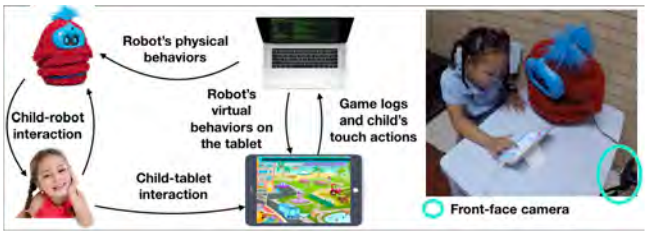

Figure 1: The integrated system for a robotic learning companion consists of a robot, a computer, a USB camera and a touch-screen tablet.

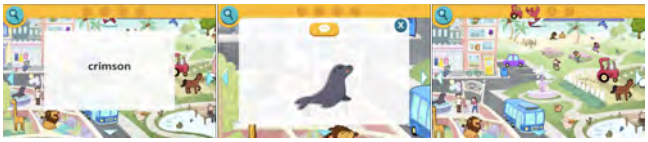

Figure 2: Overview of the *Word Quest* game. (1) A new game mission starts. (2) A game object is found, and the player can press the yellow button to check whether the object is correct. (3) Two correct game objects have been found and are shown in the top bar of the screen. When four correct objects are found, the mission is completed.

Figure 3: The *WordQuest* game sends out a quest mission, and the robot and child take turns finding objects. Each player can only find one object during their turn. When four correct objects are found irrespective of who finds them, the quest mission is completed.

3 INTERACTION DESIGN

3.1 Child-Robot Educational Game Play

Our integrated system enables a robot to interact with children and facilitate their vocabulary learning. It consists of (1) a computer hub, (2) our *Word Quest* vocabulary game on a touchscreen tablet, (3) a social robot, and (4) a suite of sensors (Figure 1). The computer hub communicates with other modules of the system using ROS (the open-source Robot Operating System).

Word Quest is a collaborative game similar to the classic game, *I Spy*, in which a child and a robot take turns identifying objects called out by a quest mission on a touchscreen tablet (Figure 2). Game missions are issued for the child and robot to complete as a team. Each game mission (e.g., "can you find objects that are in crimson?") contains one target word (e.g., "crimson") that the child needs to learn (Figure 3). The child learns its meaning by taking turns with a robot finding the objects in the game scene that represents the target word's meaning (e.g., objects that are red in color). When a candidate object is found, the player needs to press the button above the object to know whether the object is correct. Once a player receives the result, their turn is over, and it is the other player's turn. A game mission takes multiple robot-child turns to finish. When four correct objects are collected by the child-robot team, the mission is completed. The game has 11 missions in total with the following target words: *azure, gigantic, minuscule, garment, lavender, vehicle, delighted, crimson, soar, aquatic,* and *recreational activity.*

The social robot used in our study is Tega, an appealing and expressive robot designed and deployed as a learning companion for young children. To date, it has been deployed in various educational settings for different learning tasks [15, 25]. The robot is about 11 inches tall with a squash-and-stretch body with a plush exterior. The robot speaks with a child-friendly voice and can display body and facial expressions. It is able to support verbal interaction using the robot's built-in microphones and Google Automatic Speech Recognition. In our study, the robot was fully autonomous.

The sensor modules in our system collect children's interaction data from touch and vision modalities. All touch actions on the tablet screen in the game (e.g., tapping and dragging) and children's task-related data (e.g., interaction duration) are captured by the touchscreen tablet. An external USB camera located behind the tablet and aimed at the child's face is used to record the child's facial expression for affect analysis during interaction.

3.2 Tutee and Tutor Robot Companions

The robot in our word learning activity performs different sets of behaviors for each role (Table 1). In the tutor role, the robot knows the meanings of all words, behaves as a tutor that demonstrates knowledge, and gives informative feedback to the child without ever making a mistake throughout the game play. In the tutee role, the robot behaves as a novice peer who lacks the knowledge of all vocabulary words, is less competent than the child, but is eagar to learn. In this paradigm, the tutee robot has a 0.4 probability of selecting a correct object but neither explains the word's meaning nor provides explanation to the child. The intuition behind this design is that the tutee robot is only able to find correct objects by guessing and never knows the meanings of target vocabulary words. Thus, the child can only learn the word's meaning by trial-and-error while interacting with the tutee robot.

3.3 Macro and Micro Sub-Events

Our child-robot interaction contains multiple distinct task events (e.g., entire child's turn, when the robot finds an object, etc.). To more accurately evaluate how children express their affect within an episode, we extracted five levels of task events representing a wide range of time periods in the learning interaction (Figure 4).

All extracted interaction events are color-coded in Figure 4. The entire robot-child turn is the only top-level event (in pink). Then, we broke down an entire quest event into the robot's turn and child's turn to form two-player, turn-level events (in cyan and light green). The within-robot-turn/within-child-turn events consist of 8 events (in yellow), capturing key learning-related moments within either the robot's turn or the child's turn. The bottom-level events are four contingent events (in green and red) extracted from the two within-player-turn events (*robot receives result* and *child receives result*) – when the child/robot receives a correct/wrong result.

4 THE STUDY

To investigate the triadic relationship between interaction context, children's facial affect display and vocabulary learning, we designed a between-subject experiment with two conditions: *Tutor* condition in which the robot behaved as a tutor and *Tutee* condition in which the robot acted as a tutee.

4.1 Hypotheses

The main research question is that interaction contexts impact the predictive power of children's affect on their vocabulary learning.

Table 1: Robot's tutee and tutor roles with their associated behaviors.

Role	Player Turn	Robot Behaviors	Behavior Definition	Example
Tutor	Robot	Keyword Definition	Explain the meaning of the mission word	"Vehicle is something you can drive, steer or ride in."
		Game Object Selection	Always select a correct object	NA
		Vocabulary Explanation	Explain why the object the robot chooses is correct	"Train is a vehicle, because we can ride in it."
	Child	Help Offering	Offer to help the child find a correct object	"Do you need my help?"
		Keyword Definition	Explain the meaning of the mission word to the child	"Color azure means blue"
		Hint Providing	Share hints on the meaning of the word	"Azure is a color"
Tutee	Robot	Help Asking	Ask the child to help find an object	"Can you please help me find a correct object?"
		Game Object Selection	Have a 0.4 probability of selecting a correct object	NA
		Asking for Explanation	Ask why the object the robot chooses is wrong	"Can you tell me why I am wrong?"
	Child	Asking for Thoughts Process	Ask how the child finds out which object is correct	"Why did you choose this one? I want to learn from you."
		Curiosity-driven Speech	Show curiosity in what the child is going to find	"I am curious of what you will find!"

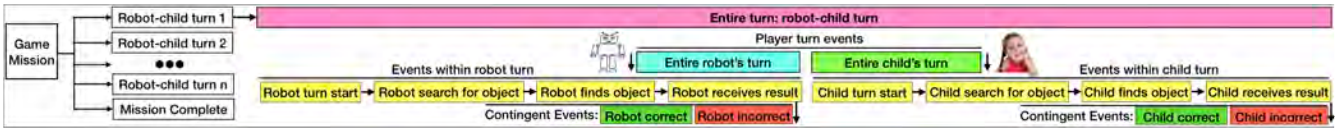

Figure 4: Given five levels of sub-events from macro to micro, 15 sub-events in total were extracted within a robot-child turn.

The interaction contexts are defined in the study as twofold: (1) social learning paradigm (H1 & H2) and (2) sub-events within an interaction (H3 & H4).

H1: Children's affective display will be significantly different between the *tutee* and *tutor* conditions.

H2: Children's affect exhibited when interacting with the *tutee* and *tutor* robots will have significantly different correlations with their vocabulary learning.

H3: Children's affect exhibited during micro-level interaction events (i.e., events occurring within either robot's turn or child's turn) will predict their vocabulary learning more strongly than their affect aggregated over macro-level interaction events (i.e., entire child's/robot's/child-robot turns).

H4: Children's affective states can be positively and negatively correlated with vocabulary learning depending on the sub-events in which they are observed. An affective state exhibited in a sub-event can be positively correlated with vocabulary growth but negatively correlated when exhibited in another sub-event.

4.2 Participants

We recruited 43 children between the ages of 5–7 from a local public school through flyers. The research was approved by our university's research ethics board. After obtaining the consent forms signed by the children's parents, we randomly divided them into the two conditions counter-balanced by age, gender, prior knowledge of target vocabulary, and English proficiency (native or English language learner). Participants were not informed of the condition they were in. A pre-test was administered to determine children's vocabulary and English proficiency levels. Two students from the *Tutee* condition and one from the *Tutor* condition withdrew from the study for reasons not related to the study (e.g, early school departure). A total of 40 children completed the study (*Tutee*: n=19, *Tutor*: n= 21). There was no statistically significant difference in children's average age (*Tutee*: 6.00 ± 0.74, *Tutor*: 5.85 ± 0.65), gender (*Tutee*: 57.89% female, *Tutor*: 61.9% female), pre-test score (*Tutee*: 2.68 ± 1.37, *Tutor*: 2.44 ± 1.43), and English proficiency (*Tutee*: 52.63% native, *Tutor*: 47.62% native) across the two study groups.

4.3 Procedure

The study protocol consisted of a pre-interaction vocabulary assessment, two *Word Quest* game sessions with the robot (one week apart), and an immediate post-interaction vocabulary assessment following each session (Fig 5). In each game session, a child played the *Word Quest* game with the robot for 20–30 minutes. In the first session, the experimenter taught the child the basics of how to play the game with the robot. The experimenter guided the child through a practice quest mission (i.e.,"can you spy something blue") for five minutes, in which the child and robot took turns finding objects in the game. After this introduction, the child and robot started the real quest missions. The experimenter stayed in the experiment room but did not intervene the learning sessions. In the *Tutee* condition, the game would automatically terminate the current mission and load the next mission if the tutee robot and child were not able to complete the mission within six minutes. This time constraint in the *Tutee* condition ensured that the child would be able to proceed and try all required game missions in a 30-minute learning session, while preventing them from struggling with a game mission for too long.

Figure 5: The experiment consisted of two learning sessions and 11 game missions in total. Two sessions were one week apart.

Figure 6: Children's vocabulary assessment administrated at both pre-test and post-test and followed the PPVT format. In this example, the examiner asks the examinee: "which one of the four stickers is a vehicle?" and then the examinee points to an object.

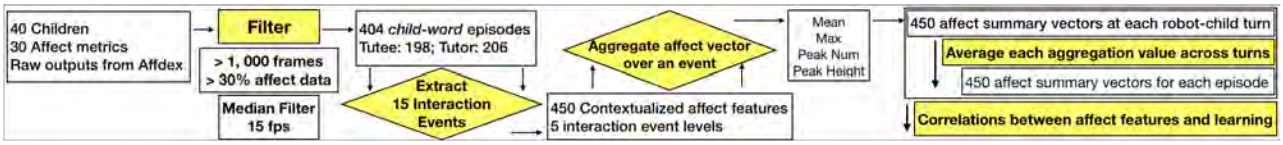

Figure 7: Data pipeline for extracting, filtering and analyzing contextualized affect features from the raw outputs of Affdex SDK. Each *child-word* episode corresponds to the time period where a child works on a quest mission with a robot.

Table 2: Mean scores and standard deviations of children's vocabulary assessments by experimental condition

Condition	Pre-Test	Post-Test	Score Change
tutee	2.68 (1.38)	4.63 (1.74)	1.95 (2.27)
tutor	2.43 (1.43)	6.00 (2.07)	3.57 (2.33)

5 EVALUATION

5.1 Children's Vocabulary Learning

5.1.1 Vocabulary assessment. The pre- and post-vocabulary tests followed the format of the Peabody Picture Vocabulary Test (PPVT) [13], in which the examinee is shown pages with four pictures on each, and the child is to point to the picture that illustrates the meaning of the stimulus word spoken by the examiner (Figure 6). To avoid random guessing and reduce false positive answers, the examiner also encouraged the child to inform the examiner if the child did not know the meaning of the stimulus word, and asked the child to explain why they selected the picture for the word. The pre- and post-tests included the 11 target words in the *Word Quest* game.

5.1.2 Vocabulary acquisition. Children's learning performance was measured by directly calculating the score difference between a child's post- and pre-vocabulary test scores. The pre-test result was used to form a baseline of participants' knowledge before the robot intervention. The immediate post-tests were administered at the end of each learning session, with words that appeared in the given session, to avoid confounding factors potentially caused by a delayed post-test (e.g., children's exposure to the test vocabulary words at school/home before the delayed post-test).

Overall, children in both tutor and tutee groups learned vocabulary words from interacting with a social robot (Table 2). Children who interacted with a tutor robot showed significantly higher learning gain, measured by the difference between pre- and post-test vocabulary scores ($t(38) = 2.224$, $p = 0.032$). Interacting with a tutor robot promoted children's learning more effectively, probably because a tutor robot's knowledge demonstration made it easier for children to grasp the meanings of target vocabulary words. In section 6, we present detailed analyses of how children's affective display correlates to their vocabulary learning for each robot role, as well as in both macro (entire turn) and micro (within specific interaction-event context) levels of the interaction.

5.2 Children's Affective Displays

The pipeline for detecting, filtering and constructing contextualized affect features for the correlation analyses is displayed in Figure 7.

5.2.1 Face-based affect detection. We video recorded children's facial expressions during learning sessions using the front-facing camera. Children's face-based affect was measured using Affdex SDK 4.0 at approximately 30 fps (a commercial tool marketed by Affectiva, Inc, Boston, MA) USA [22]. Affdex produces 30 affect-related metrics from video or images of faces. This includes estimates of seven emotions (e.g., joy, anger, contempt), 21 facial expressions (e.g., brow furrow, lip frown), one engagement metric, and one valence metric. All these metrics have a normalized range of 0 (no affective expression/ state detected) to 100 (expression or state fully present) except *Valence* that has a range of -100 (fully negative valence) to 100 (fully positive valence). For each video frame, Affdex SDK attempts to detect a face and returns a score for each affect-related metric if a face is detected. Affdex returns a null value when no face is detected.

5.2.2 Affect data filtering. After extracting real-valued affect estimates from collected videos of all children, we constructed a dataset of 440 *child-word* episodes (40 children x 11 words), each of which corresponds to a child working on a quest mission with the robot (e.g., participant #2 working on the mission *gigantic*). Then, we filtered out the episodes that have a total video frame number smaller than 1,000 (33.3 secs) or have less than 30% of its video frames containing real-valued affect data, as they lacked a substantial portion of the affect estimates due to technical failures within interaction (e.g., camera shut down during interaction or lighting was too dim to detect a face). In total, we obtained a final dataset of 404 *child-word* episodes (*Tutee*: 198 episodes, *Tutor*: 206 episodes). We applied a median filter operating over a sliding window of 15 frames (0.5*s*) to the raw real-valued affect metric vectors to smooth and filter out artifacts or dropped frames in the data.

5.2.3 affect data aggregation. To aggregate vector outputs from affect metrics into scalar values for a given time window, W, we selected four methods commonly used in signal processing: mean, max, number of peaks, and average height of detected peaks over a time window. It was implausible to find one universal aggregation method that can adequately summarize any given metric's value vector over any given event window due to the diversity of our affect features (i.e., 30 affect metrics and multiple interaction events). Max and total number of peaks were chosen to catch highly expressive yet transitory facial expressions (e.g., lip pucker) over W, as the methods are more sensitive to short signal spikes over large W (e.g., *entire robot turn*). Conversely, mean and average peak height were used to capture the overall expressivity of a metric over W.

Both the number of peaks and average height of detected peaks were measured using Scipy's built-in signal processing tool [18]. A peak is detected when the magnitude of a real-valued output from an affect metric is at least 1/4 of this metric's max value, with its prominence above 10 and a minimum distance of 5 frames between two detected peaks. For an affect metric vector during W, the four scalar values from the aggregation methods form a summary vector. Since *Valence* has a range of $[-100, 100]$, max, number of peaks and

peak height methods were applied to both its negative and positive values, returning a summary vector of 7 scalar values.

5.2.4 contextualized affect features. We measured children's affect exhibited during each sub-event within child-robot interaction extracted in Section 3.3 by applying the four affect aggregation methods (e.g., *mean*) to 30 median-filtered affect vectors (e.g., *smile*) captured during the event. Then, we obtained 30 contextualized affect features (i.e., affect summary vectors) for that particular event. Since 15 sub-events (e.g., *robot receives result*) were extracted in total, 450 contextualized affect features were generated in total. A child's smile displayed when the robot receives its result, for example, is first outputted from Affdex SDK and median-filtered as a value vector. Then, we calculated the mean, max, number of peaks and average peak height of this value vector. These four scalar values form a summary vector for smile contextualized in the event of *robot receives result*, and this four-item summary vector is one of the 450 contextualized affect features for the child.

5.3 Children's Affect-Learning Relationship

After measuring children's vocabulary acquisition in section 5.1 and extracting their contextualized affect features in section 5.2, we analyzed the relationship between each contextualized affect feature and children's vocabulary acquisition using the Kendall rank correlation. The Kendall rank correlation was selected, as most of our contextualized affect features were highly skewed according to the Levene's and Shaprio-Wilk tests. For each contextualized affect feature, the max, mean, and average peak height values in its affect summary vector were labeled as quartile bins. The quartile ranges were computed using the min and max value of the feature observed throughout the dataset in each condition. A contextualized affect feature was considered significantly correlated with learning only if at least one of its aggregation values in its summary vector demonstrated a statistical significance ($p < 0.05$), and the correlation magnitude above $|r| > 0.2$ to exclude weak correlations.

The correlation results from this analysis were presented in Table 4[1][2], and then used to evaluate the hypotheses **H2**, **H3** and **H4**.

6 RESULTS AND DISCUSSION

In this section we summarize children's overall affect display per robot role as well as the analysis results of the affect-learning correlations with respect to social learning paradigm (tutor role verses tutee role) and sub-events within child-robot interaction (micro verses macro).

6.1 Children's overall expressiveness per robot role

We measured children's affective displays during the entire learning interaction by using the four aggregation methods (mean, max, #peaks, peak height) to convert each affect metric's value vector over the two entire learning sessions into a four-item summary vector. Since the values in summary vectors for most affect metrics are significantly skewed, the Mann-Whitney u-test was used to measure, for a given affect metric's summary vector, the difference in each aggregation value between the two conditions. An affect metric was considered significantly different between the conditions only when at least one aggregation value in its summary vector has a significant u-test result ($p < 0.05$).

Table 3: Children's affect displays over the entire interaction per robot role. Children in the *Tutee* condition were significantly more expressive in 13 affective states/facial expressions except *fear*.

Affect Metric	More Expressive Condition	Aggregation Methods (U-test $p < 0.05$)
Fear	*tutor*	*peak height*
Anger	*tutee*	*max*
Contempt	*tutee*	*max*, *#peaks*
Surprise	*tutee*	*max*, *peak height*
Attention	*tutee*	*max*
Engagement	*tutee*	*#peaks*
Brow Raise	*tutee*	*max*
Eye Closure	*tutee*	*mean*, *peak height*, #peaks
Inner Brow Raise	*tutee*	*peak height*
Jaw Drop	*tutee*	*mean*, *max*, *#peaks*, *peak height*
Lid Tighten	*tutee*	*mean*, *max*, *#peaks*, *peak height*
Mouth Open	*tutee*	*mean*, *#peaks*, *peak height*
Smirk	*tutee*	*#peaks*
Upper Lip Raise	*tutee*	*max*, *peak height*

We found 14 out of 30 affect metrics significantly different by condition. Children in the *Tutee* condition were significantly more expressive than children in the *Tutor* condition for most of the 14 metrics except *fear* (Table 3). Specifically, children who interacted with the tutee robot exhibited significantly greater attention and engagement, as well as negatively-valenced affect metrics such as *contempt* and *smirk*. Therefore, the results confirmed **H1**.

Children interacting with a tutor robot received the robot's help and correct answers on every turn, so the robot's behaviors could have become predictable to children. Children may thus have become less expressive and attentive as the sessions progressed. In contrast, the tutee robot made occasional mistakes and did not find correct objects every time. The tutee robot also showed growth mindset and curiosity when finding a wrong object. This set of tutee behaviors might have elicited children's greater engagement and surprise in the game play, as well as higher attention to robot's behaviors. Similarly, children in the *Tutee* group showed greater facial expressions associated with negative valence, probably because children might have experienced more frustration and stress when learning new words without robot's guidance.

6.2 Effect of robot role on children's affect-learning correlations

As show in Table 4, 64 and 9 contextualized affect features in total were significantly correlated with vocabulary learning in the *Tutee* and *Tutor* conditions, respectively. Moreover, the *Tutee* condition had more affect metrics significantly correlated with learning in all five layers of sub-events (i.e., *entire turn, player turn, within robot turn, within child turn*, and *contingent event*) than in the *Tutor* condition.

[1]Sub-events marked with † are action-triggered instantaneous events, so the time window for these sub-events is 3 seconds capturing 1.5 seconds before and after the action occurrences.

[2]The correlation values from the affect summary vector, for most contextualized affect features listed in the table, are either positive or negative but not both, so only their largest positive/negative correlation values are displayed. Four features have both positive and negative correlations, so their largest positive and negative values are both displayed.

Table 4: Results on correlations between affect metric and learning per robot role and per sub-event.

Sub-Event Type	Sub-Event Name	Learning-Correlated Affect Metrics ($p < 0.05$, $\|r\| > 0.2$)			
		Total in *Tutee*	Total in *Tutor*	Metrics in *Tutee*	Metrics in *Tutor*
Entire Turn	Robot-child turn	6	1	anger(0.22), eyeWiden(0.2), browFurrow(0.2), noseWrinkle(0.2), chinRaise(0.21), lipPress(0.24)	engagement(-0.25)
Player Turn	Entire robot's turn	7	0	noseWrinkle(0.24), lipSuck(0.22), lipPress(0.22), valence(0.3), chinRaise(0.23), contempt(0.23), smirk(0.2)	
	Entire child's turn	2	1	lipSuck(0.21),lipPress(0.24)	lipPucker(0.24)
	All events	9	1		
Within Robot Turn	Robot turn start†	2	1	chinRaise(0.21), joy(-0.23)	valence(0.25)
	Robot search	1	0	noseWrinkle(0.21)	
	Robot finds object†	9	2	dimpler(-0.23), contempt(0.31), lipStretch(-0.21), cheekRaise(-0.21), lipSuck(-0.22), valence(-0.22), disgust(0.26), smirk(0.3), attention(-0.24)	surprise(-0.23), jawDrop(0.22)
	Robot receives result†	7	1	dimpler(-0.22), smirk:(0.28,-0.24), disgust(0.31), contempt(0.27,-0.28), upperLipRaise(-0.23), chinRaise(0.23)	attention(0.21)
	All events	19	4		
Within Child Turn	Child turn start†	1	0	dimpler(-0.22)	
	Child search	1	0	mouthOpen(-0.25)	
	Child finds object†	1	1	valence(-0.24)	eyeClosure(0.26)
	Child receives result†	3	0	valence(0.23), contempt(0.23), joy(0.22)	
	All events	6	1		
Contingent Event	Robot correct†	8	0	chinRaise(0.28), engagement(0.27), valence(-0.3), disgust(0.3), upperLipRaise(-0.24), smirk(-0.25), dimpler(-0.26), lipStretch(-0.24)	
	Robot incorrect†	4	NA	dimpler(-0.21), smirk(0.24), contempt(0.26,-0.22), disgust(0.3)	NA
	Child correct†	9	0	joy(-0.25), smile(-0.27), smirk(0.21), sadness(-0.21), noseWrinkle(-0.25), contempt(0.23,-0.25), browFurrow(-0.31), mouthOpen(-0.28), disgust(-0.22)	
	Child incorrect†	3	2	eyeWiden(0.33), negative valence(0.23), fear(0.22)	lipPucker(0.25), noseWrinkle(0.21)
	All events	24	2		
All	All events	64	9		

Table 5: Affect metrics that have both significantly positive and negative correlations with learning in different sub-events.

Affect Metric	Sub-Events Within Interaction		Robot Condition
	Max Positive Correlation ($r > 0.2$)	Max Negative Correlation ($r < -0.2$)	
brow furrow	Robot-child turn (0.205)	Child correct (-0.307)	Tutee
contempt	Robot finds object (0.306)	Robot receives correct (-0.277)	Tutee
disgust	Robot receives result (0.306)	Robot correct (-0.284)	Tutee
joy	Child receives result (0.223)	Child correct (-0.252)	Tutee
lip suck	Entire robot turn (0.221)	Robot finds object (-0.219)	Tutee
nose wrinkle	Entire robot turn (0.241)	Child correct (-0.247)	Tutee
smirk	Robot finds object (0.296)	Robot correct (-0.245)	Tutee
valence	Entire robot's turn (0.295)	Robot correct (-0.297)	Tutee

Overall, children's affect was more closely tied to their vocabulary learning when exhibited in the *Tutee* condition than in the *Tutor* condition. These results validate **H2**, and indicate that, when a tutee robot does not directly coach a child, the child's affect exhibited in the interaction becomes more crucial for their learning. When designing future educational robots that interact with children as their peers instead of tutors, it is especially important to integrate children's affect into the computational models that guide robot's behaviors and decision making.

6.3 Effect of macro and micro sub-events on affect-learning correlations

We analyzed the affect-learning correlations within each of the 15 sub-event. This provided insights into which affect features in each sub-event can lead to predicting students' vocabulary growth. Table 4 depicts the list of micro events within the robot's and child's turns ($N = 8$), and contingent events that follow each robot/child turn ($N = 4$). Our results show that the within-robot-turn events have 19 and 4 learning-correlated affect features in the *Tutee* and *Tutor* conditions, respectively. In contrast, only 6 and 1 affect features in the within-child-turn events were found significantly correlated with learning in the *Tutee* and *Tutor* conditions, respectively. Hence, for both *Tutor* and *Tutee* conditions, more affect features observed during within-robot-turn events were correlated to children's learning than during within-child-turn events. This finding indicates that children's affect exhibited in response to the robot's behaviors is a stronger predictor of their learning outcomes than their affect displayed when children themselves perform the learning task.

Among the within-robot-turn events, *Robot finds object* and *Robot receives result* have the greatest number of learning-correlated affect features in both the *Tutee* condition (*Robot finds object*: 9; *Robot receives result*: 7) and *Tutor* condition (*Robot finds object*: 2; *Robot receives result*: 1). Furthermore, children's affect expressed in the four contingent events extracted from *Robot receives result* and *Child receives result* predicted their learning outcomes best among all events from all five event layers. Specifically, 24 affective features correlated with learning in the *Tutee* group during the four contingent events, and 2 features in the *Tutor* group during the three contingent events. These results confirm **H3**.

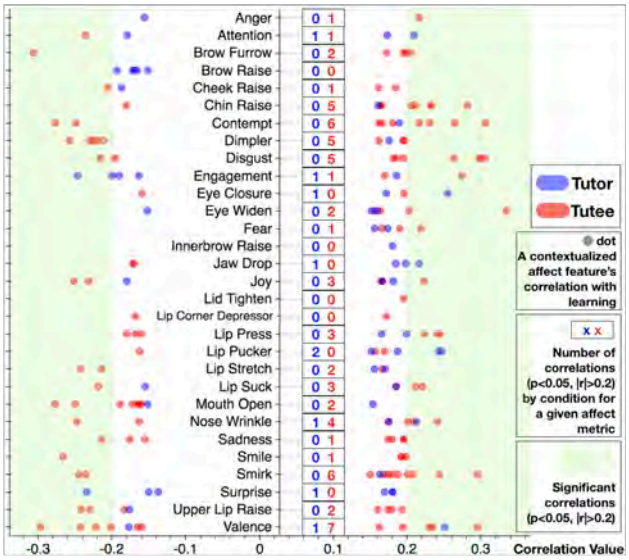

Figure 8: The correlation distribution between contextualized affect features and vocabulary learning by condition and affect metric. Multiple affect metrics have significantly positive and negative correlations with learning in different sub-events within child-robot interaction.

These results suggest that children's affective displays during micro events of small intervals within child-robot interaction can contain rich information on their learning to predict student learning outcomes. In contrast, aggregating affect exhibited throughout the entire learning interaction does not necessarily lead to accurate student learning prediction. Thus, when designing future affect-aware pedagogical agents, it can be more effective and efficient to capture student affect exhibited during short critical time periods within the interaction (e.g., when either robot or child receives feedback on their learning attempts). Last, the difference in the number of significant features during contingent events between two conditions (*Tutor*: $N = 2$; *Tutee*: $N = 24$) may explain why children's affect exhibited when interacting with a *tutee* was a strong predictor of their learning outcomes. Namely, children's affective response to the results of the robot's and their own attempts helped them figure out the meanings of quest words when the robot did not directly give them any hints or help.

6.4 Affect metrics having both positive and negative correlations in different sub-events

Among 30 affect metrics, 8 metrics in the *Tutee* condition have both significantly positive and negative correlations with children's learning ($p < 0.05$; $|r| > 0.2$) in different sub-events (Figure 8, Table 5). None of the metrics from the *Tutor* condition have both significantly positive and negative correlations. *Brow furrow*, for example, shows a wide range of correlations ($r \in [-0.307, 0.205]$). It is positively correlated with learning when exhibited during the *entire-robot-child-turn* event ($r = 0.205$), while negatively correlated with learning when the child receives the result of their correct attempt ($r = -0.307$). These results show that the relation between children's affective displays and learning is neither fixed nor unidirectional, but rather is contingent on the interaction events.

Furthermore, some negatively-valenced affect metrics including *contempt*, *disgust* and *smirk* were positively correlated with children's learning outcomes when exhibited during either the *robot-find-object* or *robot-receive-result* events in the *Tutee* condition (*contempt*: $r = 0.306$; *disgust*: $r = 0.306$, *smirk*: $r = 0.296$). Admittedly, one should not interpret the negatively-valenced affect labels at face value. These active expressions, however, may indicate that children were highly engaged in evaluating and affectively reacting to the robot's learning performance. This active engagement helped them more accurately infer about word meanings through the tutee robot's trial-and-error. This finding further strengthens the potential of negative affect as facilitator of learning in certain contexts, and resonates with the prior research showing that negative affects including confusion and sadness improved learning under certain situations [12, 24].

Overall, these results show that children's affective displays combined with associated interaction contexts are correlated with word learning, in contrast to just considering children's affective displays alone, confirming **H4**. When designing future affect-aware pedagogical agents, the correlation direction (positive/negative) between student affect and learning outcomes needs to be contextualized in specific sub-events within child-robot interaction, rather than pre-determined, as an affect metric may facilitate or inhibit learning when displayed in different micro- and macro-level interaction events.

7 CONCLUSION AND FUTURE WORK

Designing educational robots that can successfully leverage student facial affect to promote student learning poses a significant challenge, namely, how to understand the complex relationships between interaction context, student affect, and learning. We collected a rich dataset to perform a detailed analysis of the correlations between these three factors. We showed that both robot role and sub-events within student-robot interaction modulate the relationship between student affect and learning outcomes.

The affect-learning relationship for some contextualized affect features was more sensitive to the four affect aggregation methods, as we found that 4 of 73 contextualized affect features had both significantly positive and negative correlations returned by the aggregation methods (Table. 4). Thus, we plan to analyze how different affect aggregation methods impact the affect-learning correlations. Second, we plan to use different commercial affect extraction tools (e.g., FACET) to capture children's facial affect and compare their detection accuracy, as they have only been evaluated on datasets of adults' faces [34]. Given the insights in this paper, we also plan to integrate children's affect into state-of-art student cognitive models (e.g.,[16]) to foster children's vocabulary learning in child-robot interaction.

8 ACKNOWLEDGMENT

This work was supported by NRI IIS 1734443, NSF IIP-1717362 and Intel Graduate Fellowship. We sincerely thank Shang-Yun Wu for developing game assets and helping run the study.

REFERENCES

[1] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (2018). https://doi.org/10.1126/scirobotics.aat5954

[2] Herbert Bless, Norbert Schwarz, Gerald Clore, Verena Golisano, Christina Rabe, and Marcus Wolk. 1996. Mood and the Use of Scripts: Does a Happy Mood Really Lead to Mindlessness? *Journal of personality and social psychology* 71 (1996), 665–79. https://doi.org/10.1037/0022-3514.71.4.665

[3] Lavonda Brown and Ayanna M. Howard. 2015. Engaging children in math education using a socially interactive humanoid robot. In *IEEE-RAS International Conference on Humanoid Robots*. https://doi.org/10.1109/HUMANOIDS.2013.7029974

[4] Rafael A Calvo and Sidney K D'Mello. 2011. *New perspectives on affect and learning technologies*. Vol. 3. Springer Science & Business Media.

[5] Ginevra Castellano, Iolanda Leite, André Pereira, Carlos Martinho, Ana Paiva, and Peter Mcowan. 2013. Multimodal affect modeling and recognition for empathic robot companions. *International Journal of Humanoid Robotics* 10 (03 2013). https://doi.org/10.1142/S0219843613500102

[6] Ginevra Castellano, Iolanda Leite, André Pereira, Carlos Martinho, Ana Paiva, and Peter W Mcowan. 2014. Context-sensitive affect recognition for a robotic game companion. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 4, 2 (2014).

[7] Huili Chen, Hae Won Park, and Cynthia Breazeal. in review. Teaching and Learning with Children: Impact of Reciprocal Peer Learning with a Social Robot on Children's Learning and Emotive Engagement. *Computers & Education* (in review).

[8] C. Clabaugh, G. Ragusa, F. Sha, and M. Mataric. 2015. Designing a socially assistive robot for personalized number concepts learning in preschool children. In *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. 314–319.

[9] Gerald Coles. 1998. *Reading lessons: The debate over literacy*. Hill & Wang. x, 212–x, 212 pages.

[10] Cristina Conati and Heather Maclaren. 2005. Data-Driven Refinement of a Probabilistic Model of User Affect. In *User Modeling 2005*. Springer Berlin Heidelberg, 40–49.

[11] Fernando De la Torre and Jeffrey F Cohn. 2011. Facial expression analysis. In *Visual analysis of humans*. Springer, 377–409.

[12] Sidney D'Mello, Blair Lehman, Reinhard Pekrun, and Art Graesser. 2014. Confusion can be beneficial for learning. *Learning and Instruction* 29 (2014), 153 – 170.

[13] Lloyd M Dunn and Douglas M Dunn. 2007. *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

[14] Amir Erez and Alice M Isen. 2003. The influence of positive affect on the components of expectancy motivation. *The Journal of applied psychology* 87 (2003), 1055–67. https://doi.org/10.1037//0021-9010.87.6.1055

[15] Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, J J Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and C Breazeal. 2016. Affective personalization of a social robot tutor for children's second language skills. *Proceedings of the 30th Conference on Artificial Intelligence* (2016).

[16] Ishaan Grover, Hae Won Park, and Cynthia Breazeal. 2019. A Semantics-based Model for Predicting Children's Vocabulary. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*.

[17] M. E. Hoque, D. J. McDuff, and R. W. Picard. 2012. Exploring Temporal Patterns in Classifying Frustrated and Delighted Smiles. *IEEE Transactions on Affective Computing* 3, 3 (2012), 323–334. https://doi.org/10.1109/T-AFFC.2012.11

[18] Eric Jones, Travis Oliphant, Pearu Peterson, et al. 2001–. SciPy: Open source scientific tools for Python. http://www.scipy.org/

[19] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive robots as social partners and peer tutors for children : A field trial. *Human-Computer Interaction* (2004). https://doi.org/10.1207/s15327051hci1901&2_4

[20] James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. 2016. Social robot tutoring for child second language learning. In *ACM/IEEE International Conference on Human-Robot Interaction*. https://doi.org/10.1109/HRI.2016.7451757

[21] Tsuyoshi Komatsubara, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2014. Can a Social Robot Help Children's Understanding of Science in Classrooms?. In *Proceedings of the Second International Conference on Human-agent Interaction (HAI '14)*. ACM, New York, NY, USA, 83–90. https://doi.org/10.1145/2658861.2658881

[22] Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. In *Proceedings of the 2016 CHI Conference Extended Abstracts*. ACM.

[23] Debra K Meyer and Julianne C Turner. 2006. Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review* 18, 4 (2006), 377–390.

[24] Caitlin Mills, Jennifer Wu, and Sidney D'Mello. 2019. Being Sad Is Not Always Bad: The Influence of Affect on Expository Text Comprehension. *Discourse Processes* 56, 2 (2019), 99–116. https://doi.org/10.1080/0163853X.2017.1381059 arXiv:https://doi.org/10.1080/0163853X.2017.1381059

[25] Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. 2017. Telling stories to robots: The effect of backchanneling on a child's storytelling. In *ACM/IEEE International Conference on Human-Robot Interaction*.

[26] Hae Won Park, Ishaan Grover, Samuel Spaulding, Louis Gomez, and Cynthia Breazeal. 2019. A Model-free Affective Reinforcement Learning Approach to Personalization of an Autonomous Social Robot Companion for Early Literacy Education. *Proceedings of the 33th Conference on Artificial Intelligence* (2019).

[27] Reinhard Pekrun. 2006. The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review* 18 (2006), 315–341. https://doi.org/10.1007/s10648-006-9029-9

[28] Reinhard Pekrun, Thomas Goetz, Wolfram Titz, and Raymond P. Perry. 2002. Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist* 37, 2 (2002), 91–105. https://doi.org/10.1207/S15326985EP3702_4

[29] Rajagopal Raghunathan and Yaacov Trope. 2002. Walking the tightrope between feeling good and being accurate: mood as a resource in processing persuasive messages. *Journal of personality and social psychology* 83 3 (2002), 510–25.

[30] T. Schodde, K. Bergmann, and S. Kopp. 2017. Adaptive Robot Language Tutoring Based on Bayesian Knowledge Tracing and Predictive Decision-Making. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. 128–136.

[31] S. Spaulding and C. Breazeal. 2019. Frustratingly Easy Personalization for Real-Time Affect Interpretation of Facial Expression. *International Conference on Affective Computing and Intelligent Interaction and workshops* (Sep 2019). http://par.nsf.gov/biblio/10108260

[32] Samuel Spaulding, Huili Chen, Safinah Ali, Michael Kulinski, and Cynthia Breazeal. 2018. A Social Robot System for Modeling Children's Word Pronunciation: Socially Interactive Agents Track. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 1658–1666.

[33] Samuel Spaulding, Goren Gordon, and Cynthia Breazeal. 2016. Affect-Aware Student Models for Robot Tutors. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. 864–872.

[34] Sabrina Stöckli, Michael Schulte-Mecklenbeck, Stefan Borer, and Andrea C. Samson. 2018. Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods* 50, 4 (01 Aug 2018), 1446–1460. https://doi.org/10.3758/s13428-017-0996-1

[35] Fumihide Tanaka and Shizuko Matsuzoe. 2012. Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning. *Journal of Human-Robot Interaction* (2012). https://doi.org/10.5898/JHRI.1.1.Tanaka

[36] Susanne Vogel and Lars Schwabe. 2016. Learning and memory under stress: implications for the classroom. In *Science of Learning*.

[37] Beverly Woolf, Winslow Burleson, Ivon Arroyo, Toby Dragon, David Cooper, and Rosalind Picard. 2009. Affect-aware tutors: Recognizing and responding to student affect. *IJLT* 4 (01 2009), 129–164. https://doi.org/10.1504/IJLT.2009.028804