The Affective Impact of Tutor Questions: Predicting Frustration and Engagement

Alexandria K. Vail
Department of Computer Science
North Carolina State University
Raleigh, North Carolina
akvail@ncsu.edu

Joseph B. Wiggins
Department of Computer Science
North Carolina State University
Raleigh, North Carolina
jbwiggi3@ncsu.edu

Joseph F. Grafsgaard
Department of Psychology
North Carolina State University
Raleigh, North Carolina
jfgrafsg@ncsu.edu

Kristy Elizabeth Boyer
Department of Computer & Information Science & Engineering
University of Florida
Gainesville, Florida
keboyer@ufl.edu

Eric N. Wiebe
Department of STEM Education
North Carolina State University
Raleigh, North Carolina
wiebe@ncsu.edu

James C. Lester
Department of Computer Science
North Carolina State University
Raleigh, North Carolina
lester@ncsu.edu

ABSTRACT

Tutorial dialogue is a highly effective way to support student learning. It is widely recognized that tutor dialogue moves can significantly influence learning outcomes, but the ways in which tutor moves, student affective response, and outcomes are related remains an open question. This paper presents an analysis of student affective response, as evidenced by multimodal data streams, immediately following tutor questions. The findings suggest that students' affect immediately following tutor questions is highly predictive of end-of-session self-reported engagement and frustration. Notably, facial action units which have been associated with emotional states such as embarrassment, disgust, and happiness appear to play important roles in students' expressions of frustration and engagement during learning. This line of investigation will aid in the development of a deeper understanding of the relationships between tutorial dialogue and student affect during learning.

Keywords

Tutorial dialogue, affect, frustration, engagement, facial expression

1. INTRODUCTION

Tutorial dialogue provides rich, natural language adaptation to students during learning. An understanding has emerged about the role of interactivity in tutorial dialogue [40, 6] and on dialogue strategies for most effectively supporting students in task-oriented tutorial dialogues [29, 10]. However, a pressing issue is developing an understanding of how specific tutor dialogue moves impact students' affect, and in turn, what influence students' affective responses may have on outcomes.

The need for modeling affect during learning is widely recognized. Research has shown that suites of affect detectors from sensors and log files can perform well but that there are trade-offs depending on the goals of the affect detection modules [22, 33]. Affect detectors have been investigated for a wide variety of affective states including confidence, excitement, frustration, and interest [41], and within tutorial dialogue, for uncertainty [11]. There have also been great strides in sensor-free affect detection which relies primarily on log files [2]. This approach has shown promise during cognitive tutoring [9] and for distinguishing frustration and confusion [27].

Out of all of the affective phenomena that have been examined during learning, two affective states are frustration and engagement. These states have been examined in fine-grained analyses as tutoring unfolds, and also as outcome measures regarding students' perceptions of the success of the tutoring session. Engagement and frustration have been predicted at above-chance levels using facial expression-based affect detection even without the presence of interactive events during text or diagram comprehension [5]. Engagement and frustration have also been predicted with nonverbal behaviors, including facial expression, after student task events during problem solving [16]. In a compelling development, emerging evidence shows that fine-grained affective events can have long-lasting relationships with outcomes that may be far removed from those affective events [36].

This paper advances the understanding of student emotions in learning by examining students' fine-grained affective responses to tutor questions during tutorial dialogue. It investigates the hypothesis that students' affective responses immediately following tutor questions are related to self-reported frustration and engagement at the end of the session. The results indicate that several key facial expression

DOI: 10.1145/1235

features immediately following two different types of tutor questions are highly predictive of end-of-session self-reported engagement and frustration. This line of investigation represents a step forward in understanding the affective impact of tutorial strategies.

2. RELATED WORK

Tutorial dialogue researchers have long studied what human tutors naturally do: how strategies differ between experts and novice tutors [12] whether Socratic or didactic approaches are most effective [35] and how tutors scaffold and fade support during problem solving [4], among others. The impact of particular tutorial dialogue moves has been the focus of significant attention, with findings indicating that positive and negative feedback have different impact based on students' self-efficacy level [3], that bottom-out directives are not conducive to learning [29], and that adapting to student uncertainty improves the effectiveness of tutorial dialogue [10]. However, this paper examines a different aspect of these tutorial dialogue moves that is critical in learning: students' affective response as expressed on the face and as embodied in gestures.

Multimodal features such as dialogue, facial expression, posture, and task actions have been used to predict affective states, such as boredom, confusion, excitement, and frustration, as those states occur during learning [23, 8, 7]. Moreover, multimodal features such as facial expression and gestures can significantly predict frustration and engagement reported at the end of tutoring sessions [17], and some differences have emerged in the extent to which upper and lower facial expression features are associated with these outcomes [15]. This previous work on utilizing multimodal features for predicting frustration and engagement during human-human tutoring has emphasized the important role that tutor dialogue moves play in affective outcomes. Other factors, such as student personality profile, can also contribute significantly to predicting these outcomes [39]. The present work examines moment-by-moment affect as evidenced by multimodal traces, and then analyzes the relationship between these multimodal behaviors and the outcomes of frustration and engagement as reported by students after the tutoring session.

3. STUDY DATA

The present analysis investigates the multimodal behavior of students during a computer-mediated tutorial session in introductory computer science, and specifically in Java programming [18, 30]. The tutorial interface, shown in Figure 1, is divided into four panes: the task description, the student's Java source code, the compilation and execution output of the program, and the textual dialogue messages between the tutor and the student. The tutor's interactions with the environment were constrained to progression between tasks and sending textual messages to the student.

Students ($N = 67$) were university students in the United States enrolled in an introductory engineering course, with an average age of 18.5 years ($s = 1.5$ years), whereas the human tutors ($N = 5$) were primarily graduate students with previous experience in tutoring or teaching introductory programming. The behavior of the student was collected using a set of multimodal sensors, as shown in Figure 2, including a Kinect depth sensor, an integrated webcam, and a skin conductance bracelet. The following subsections detail the modalities appearing significant in the present analysis.

Each student participated in six 40-minute sessions over the course of four weeks; however, the present analysis only examines data from the first lesson. Before and after each lesson, students completed a content-based pretest and identical posttest; the tutoring sessions were found to be significantly effective in facilitating learning gains ($p \ll 0.0001$). In addition to the posttest, students also completed a post-survey, including the NASA-TLX workload survey [20] and the User Engagement Survey [32]. The present analysis investigates self-reported *frustration*, taken from the Frustration Level item of the NASA-TLX workload survey, and *engagement*, taken as an average of three sub-scales of the User Engagement Survey: Focused Attention (perception of time passing), Felt Involvement (perception of involvement with the session), and Endurability (perception of the activity as worthwhile).

3.1 Task Event and Dialogue Features

During the tutoring session, the interface described above logged tutor and student dialogue messages, student typing in the code window, and student progress through the task. No turn-taking measures were enforced in the dialogue: students and tutors could send messages to the other at any point. All exchanged messages were automatically tagged by a J48 decision tree classifier [37] with a dialogue act annotation scheme created for task-oriented tutorial dialogue that differentiates tutor questions, feedback, and hints, among other dialogue moves [38]. In that work, the Cohen's kappa between two human annotators was 0.87 and the Cohen's kappa between human and the J48 decision tree classifier was 0.786.

The analysis presented here focuses on two types of tutor dialogue moves: inference questions and evaluative questions. (Although other question types were investigated, student reactions to these were not found to have significant predictive power.) Inference questions require the formation of an action plan or reasoning about existing content knowledge. For example, *'How do you think this problem can be solved?'*, or *'How can you fix this error?'* are considered to be inference questions. On the other hand, evaluative questions aim to evaluate the student's belief in his or her own understanding of the material, e.g., *'Does that make sense so far?'*, or *'Do you understand?'* (see Figure 4).

Previous work has suggested that questions can stimulate cognitive disequilibrium in a student [34], which is often considered to be a critical step in knowledge acquisition [13]. On the other hand, evaluative questions that ask a novice to evaluate whether she understands material may not be particularly helpful pedagogically because novices often cannot identify what they do not understand, or may be hesitant to speak up even if they are aware that they are confused. Nonetheless these questions occurred regularly in our corpus with experienced (though not expert) human tutors. We investigate whether students' affective response to these types of tutor dialogue moves is significantly predictive of student engagement and frustration as reported at the end of the session.

3.2 Facial Expression Features

Student facial expressions were automatically extracted

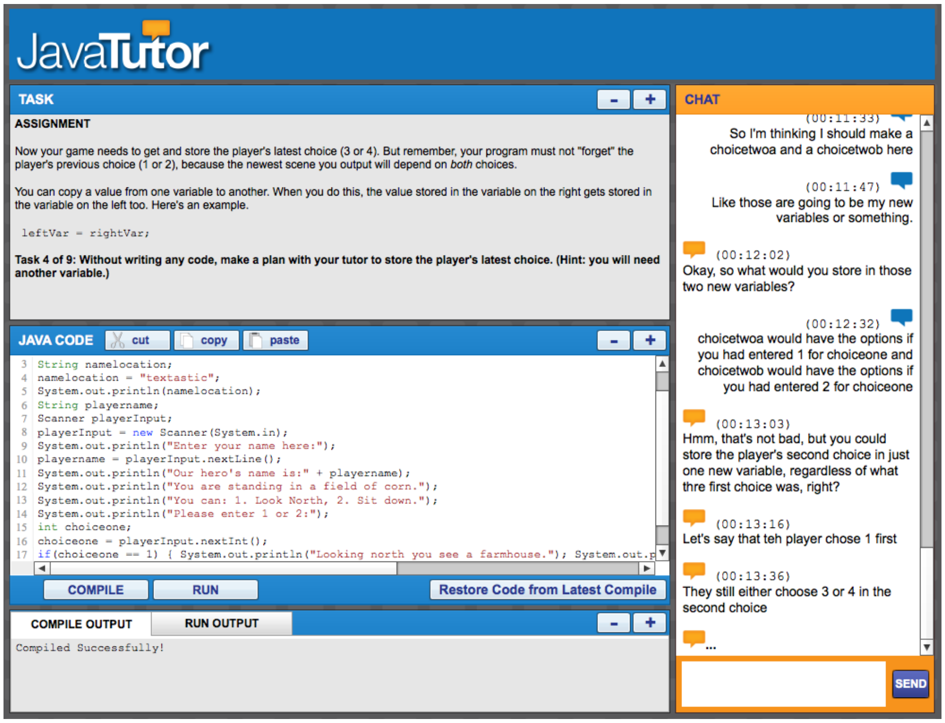


Figure 1: The web-based tutorial interface for Java programming.

using a state-of-the-art facial expression recognition toolbox, FACET (commercial software preceded by a research version known as the Computer Emotion Recognition Toolbox, CERT) [26]. FACET tracks the frame-by-frame presence of several facial action units according to the Facial Action Coding Scheme [25]. These action units include movements such as AU6 Cheek Raiser, AU12 Lip Corner Puller, AU24 Lip Pressor, and AU26 Jaw Drop (see Figures 5 and 6 for illustration). For each facial action unit, the FACET software suggests an *Evidence* measure, indicating the chance that the target expression is present. This *Evidence* measure is on a scale where negative values represent evidence of the absence of a facial expression and positive values indicate evidence of the presence of one. The more positive the measure, the more confident FACET is that the feature is present.

3.3 Gesture Features

The Kinect depth camera also tracked hand-to-face gestures made by the student during the tutoring session. An algorithm developed to detect such gestures was developed to recognize one or two hands touching the lower face. In order to do this, the algorithm relies on surface propagation from the center of the head, identifying round (i.e. a normal head shape) or oblong shapes (i.e., shapes extending beyond the normal head shape) based on distances from the center of the head. This gesture detection algorithm was previously found to be 92.6% accurate when compared against manual labels [14].

4. ANALYSIS

The present analysis focuses on the affective response of a student, as observed by multimodal traces of face and gesture, after tutor inference questions and evaluative questions. We hypothesize that multimodal features after these tutor questions can predict student engagement and frustration. In particular, we examine three seconds after each tutor dialogue move (a manually-determined interval). The multimodal response of the student was characterized using the following categories of features, all of which were provided to the predictive models. However, note that only the first two of these categories of features (shown in bold below) appear significantly predictive within the models.

1. **Average evidence measure for each of the facial expression action units during the interval (19 features)**

2. **Percentage of the interval in which a one-hand-to-face or two-hands-to-face gesture was observed (2 features)**

3. Number of skin conductance responses identified during the interval as measured by a skin conductance response bracelet (1 feature)

4. Average student distance from the workstation during the interval (1 feature)

5. Average difference between the highest and lowest points of the student's body from the workstation during the interval, indicating leaning (1 feature)

We calculated the average value of each multimodal feature listed in the categories above across each tutoring session. For each feature, we computed its conditional probability of occurring after the tutor moves of inference question or evaluative question. We also provided the model with the overall occurrence of that feature across the entire tutoring

Figure 2: Multimodal instrumented tutoring session, including a Kinect depth camera to detect posture and gesture, a webcam to detect facial expression changes, and a skin conductance bracelet to detect electrodermal activity.

Figure 3: Dialogue excerpt illustrating a tutor inference question in context.

Student compiles the program, encounters an error.

STUDENT	Oh.
TUTOR	So how can we fix this?
STUDENT	Hmm.
STUDENT	Switch the prompt line with the response line?
TUTOR	Okay, try it.

Figure 4: Dialogue excerpt illustrating a tutor evaluative question in context.

STUDENT	Do I need to set the player input before line 13?
TUTOR	The `while` tests that [variable]. You need to be sure it enters the loop at least once.
TUTOR	Good.
TUTOR	Does that make sense?
STUDENT	Yeah.
STUDENT	But what happens if I don't enter 1 or 2?

session in order to control for the influence of the feature overall (rather than only after the tutor moves of interest). Specifically, the features conditional on tutor moves were averages of the form $Avg(Feature|TutorQ)$ for each student that completed the session. The session-wide average of each feature, $Avg(Feature)$ were also provided to the model for each multimodal feature in all of the categories above.

Standardization was performed on each feature by subtracting the mean and dividing by the standard deviation, so that the regression coefficients would be more interpretable. The standardized features were provided to a stepwise regression modeling procedure optimizing for the leave-one-student-out cross-validated R^2 value (the coefficient of determination), while at the same time requiring a strict $p < 0.05$ cut-off value after Bonferroni correction on significance values.

5. RESULTS AND DISCUSSION

For both types of tutor question, evaluative and inference, a predictive model was built to predict student frustration and student engagement, resulting in a potential four models. Three of the four models uncovered significant predictive relationships. The following subsections detail models predicting frustration after tutor inference and evaluative questions, and a model predicting engagement after tutor evaluative questions.

5.1 Frustration

The results suggest that student facial expressions are significantly predictive of self-reported end-of-session frustration. The predictive model for student frustration based on tutor evaluative questions includes two features, both of which are facial action units occurring in the three-second interval following the tutor evaluative question (Table 1).

Two facial action unit features after tutor evaluative ques-

[1]The models reported in this paper were built as a part of a larger exploratory analysis. As a result, the p-values reported have been modified by a Bonferroni correction

Table 1: Predictive model for standardized end-of-session frustration after tutor evaluative questions (TutorQE).[1]

Frustration =	R^2	p
-0.7039 * AU12 after TutorQE	0.0764	0.014
-0.6279 * AU28 after TutorQE	0.2471	0.030
-0.1635 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.3235$		

Table 2: Predictive model for standardized end-of-session frustration after tutor inference questions.[1]

Frustration =	R^2	p
$+0.5660$ * AU6 after TutorIQ	0.2893	0.022
$+0.3635$ * AU20	0.0499	0.019
-0.0174 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.3392$		

tions are significantly predictive of student frustration. Higher intensity levels of of AU12 Lip Corner Puller (Figure 5b) following a tutor evaluative question are *negatively* indicative of frustration, as is the presence of AU28 Lip Suck (Figure 5d). AU12 is associated with smiling, which is typically not associated with frustration although on occasion, the two can go hand in hand [21].

AU 28 is a type of lower face movement sometimes associated with fidgeting, and this type of motion may be a "self-manipulator" that is part of emotion regulation. It is possible that students engaged in this challenging learning task may exhibit this movement to alleviate negative emotions related to frustration, resulting in lower self-reported frustration at the end of the session. When students are faced with a question that asks them to evaluate whether they understand the material being tutored, these facial expressions may both reflect the presence of emotion regulation that could mitigate the students' overall feeling of frustration.

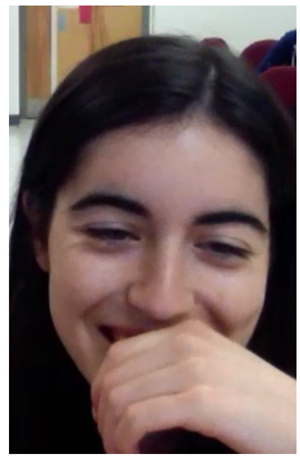
The next model examines student responses to tutor inference questions. In contrast to evaluative questions, inference questions ask students to bring pieces of knowledge together to infer the answer to a question and then to express a substantive answer. Two facial action unit features exhibited following these questions appear as significantly predictive of student frustration. The model shows that AU6 Cheek Raiser (Figure 5a) after tutor inference questions is positively predictive of frustration, as is the overall session occurrence of AU20 Lip Stretcher (Figure 5c). The model is displayed in Table 2.

Interestingly, AU6 has been related to pain expressions in the literature on pain detection [28]. When asked to answer an inference question, it is possible that students exhibited a "pained" expression that coincides with frustration. The expression of AU20 has been observed to coincide with moments of embarrassment or awkwardness [24], when people were embarrassed or amused in the period after doing directed facial actions (the technique used to develop images for the Facial Action Coding System). AU20 only occurred among embarrassed participants in that study. When faced with a tutor inference question, this expression may indicate that the student is unsure, awkward, or embarrassed, which may unsurprisingly be related to frustration. Deeper future investigation of subsequent student dialogue moves will help elucidate this phenomenon.

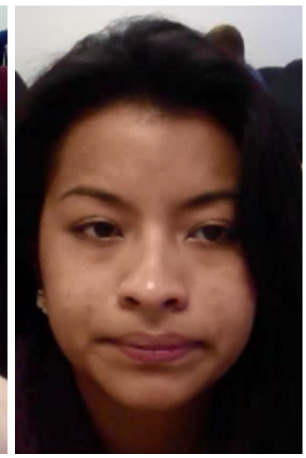
5.2 Engagement

Next we built models to predict student engagement based on affective responses to tutor inference questions and evaluative questions. For inference questions, none of the features provided to the model were predictive of engagement. However, for affective response to tutor evaluative questions, there were seven predictive features, three of which are specific to the interval following the event, and four of which are session-wide (Table 3).

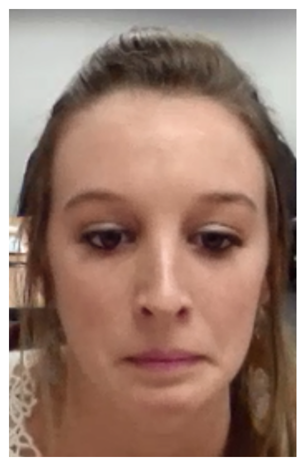
The model suggests that facial expression features account for most of the variance in predicting student engagement; however, one session-wide gesture feature was also



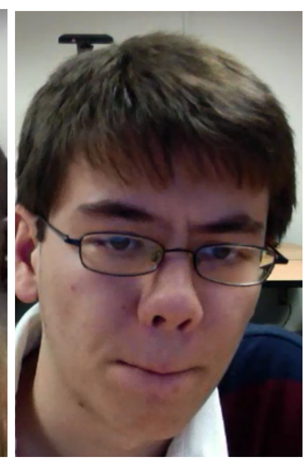
(a) AU6
Cheek Raiser + One Hand to Face



(b) AU12
Lip Corner Puller



(c) AU20
Lip Stretcher



(d) AU28
Lip Suck

Figure 5: Sample frames from the student webcam illustrating the facial action unit features appearing in the predictive models for student frustration, as identified by FACET.

$p \leq \alpha/n$, where $n = 21$ is the number of statistical tests conducted in the larger analysis, in order to reduce the familywise error rate to $\alpha = 0.05$.

Table 3: Predictive model for standardized engagement after tutor evaluative questions.[1]

Engagement =	R^2	p
+0.4422 * OneHTF	0.1815	< 0.001
−0.5989 * AU10 after TutorEQ	0.1831	< 0.001
+0.5770 * AU12	0.2280	< 0.001
+0.5097 * AU26 after TutorEQ	0.0514	< 0.001
−0.2941 * AU2	0.1923	0.003
+0.2467 * AU5	0.0295	0.002
+0.1792 * AU24 after TutorEQ	0.0566	0.018
+0.4100 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.9224$		

selected. The more frequently a student was displaying a OneHandToFace gesture, which may indicate thoughtful contemplation, the more engaging the student reported the experience at the end of the session.

Three more session-wide facial expression features were selected as significantly predictive of student engagement. The more intense the expression of AU12 Lip Corner Puller (Figure 5b) or AU5 Upper Lid Raiser (Figure 6b), the more engaged the student. For AU12 which is often associated with smiling, a positive emotion is likely related to higher engagement. In this task, AU5 is likely associated with the student looking at the screen, possibly indicating paying attention and focusing on the task (as opposed to the opposite facial movement of blinking or shutting one's eyes). In contrast, AU2 Outer Brow Raiser (Figure 6a) was predictive of lower engagement. This action unit is a component of the "fear brow" (AU1+2+4) which has been evidenced as a display of anxiety [19].

Narrowing down to the context of three seconds after tutor evaluative questions, three facial expression features were significantly correlated with student engagement. The more that a student expresses AU26 Jaw Drop (Figure 6e), or the more that the student expresses AU24 Lip Pressor (Figure 6d), the more engaged the student reported being at the end of the session. Jaw drop is a dynamic action unit that may occur when the mouth is closed or already partly open. In either case, this action unit may be associated with focus on the task, although it could also plausibly be associated with a yawn (which we would not expect to coincide with higher engagement). With respect to AU24, which is a prototypical component of anger, an important interplay of learning and affect expression emerges. Some facial movements that are part of prototypical displays of negative basic emotions, such as anger, appear to be indicative of mental effort during learning, rather than negative affect [31]. From this perspective, it makes sense that this AU24 would be related to engagement. On the other hand, the more that a student expressed AU10 Upper Lip Raiser (Figure 6c) during this interval, the *less* engagement reported by the student at the end of the session. This action unit, which is a component of prototypical disgust, is likely to run contrary to engagement.

6. CONCLUSION

Tutor dialogue moves in one-on-one human tutoring significantly influence student outcomes, both cognitive and affective. This paper has examined students' affective response to two types of tutor questions: inference questions which require some reasoning to construct an answer, and evaluative questions, which ask students to reflect on the extent to which they understand the material. The results show that immediately after these tutor questions, students' affective displays—particularly with respect to facial expression—are highly predictive of the outcomes of frustration and engagement. By detecting these affective displays which have been associated in prior studies with emotions such as embarrassment, disgust, or happiness, we can begin to understand the moment-by-moment affective processes that influence learning through tutorial dialogue, and relate those fine-grained events to overall outcomes.

While these facial movements have been associated with prototypical emotion displays in the literature, it is important to further contextualize the moments in which these expressions appear during tutoring. For instance, action units typically associated with anger are likely indicators of mental effort during learning. Similarly, an action unit associated with disgust (e.g., AU10) may be related to students' appraisal of the tutor's question in the moment. Further research seek to ground these interpretations more extensively across salient moments of tutoring.

There are several additional directions for future work. Detecting important moments during tutoring is an open area of investigation, with evidence suggesting that moment-by-moment affect may be related to distal outcomes [36, 1]. In future work, it will be important to expand our understanding of the identified non-verbal predictors for frustration and engagement more deeply. We must consider a wider variety of contexts, and explore different widths of time after tutorial events to examine affective responses with longer (or shorter) times to manifest. It is hoped that this line of investigation will lead to richer affect models for tutorial dialogue.

Acknowledgments

The authors wish to thank the members of the LearnDialogue and Intellimedia groups at North Carolina State University for their helpful input. This work is supported in part by the Department of Computer Science at North Carolina State University and the National Science Foundation through Grants IIS-1409639, CNS-1453520, and a Graduate Research Fellowship. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

7. REFERENCES

[1] R. S. Baker, A. B. Goldstein, and N. T. Heffernan. Detecting learning moment-by-moment. *International Journal of Artificial Intelligence in Education*, 21(1-2):5–25, 2011.

[2] R. S. Baker, S. M. Gowda, M. Wixon, J. Kalka, A. Z. Wagner, A. Salvi, V. Aleven, G. W. Kusbit, J. Ocumpaugh, and L. Rossi. Towards sensor-free affect detection in cognitive tutor algebra. In *Proceedings of the 5th International Conference on Educational Data Mining (EDM 2012)*, pages 126–133, 2012.

[3] K. E. Boyer, R. Phillips, M. Wallis, M. Vouk, and J. Lester. Balancing cognitive and motivational scaffolding in tutorial dialogue. In *Intelligent Tutoring Systems*, pages 239–249. Springer, 2008.

[4] W. L. Cade, J. L. Copeland, N. K. Person, and S. K. D'Mello. Dialogue modes in expert tutoring. In *Intelligent tutoring systems*, pages 470–479. Springer, 2008.

[5] Y. Chen, N. Bosch, and S. D'Mello. Video-based affect detection in noninteractive learning environments.

[6] M. T. Chi, S. A. Siler, H. Jeong, T. Yamauchi, and R. G. Hausmann. Learning from human tutoring. *Cognitive Science*, 25(4):471–533, 2001.

[7] D. G. Cooper, K. Muldner, I. Arroyo, B. P. Woolf, and W. Burleson. Ranking feature sets for emotion models used in classroom based intelligent tutoring systems. In *User Modeling, Adaptation, and Personalization*, pages 135–146. Springer, 2010.

[8] S. K. 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(2):147–187, 2010.

[9] S. E. Fancsali. Causal discovery with models: behavior, affect, and learning in cognitive tutor algebra. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014)*, pages 28–35. Citeseer, 2014.

[10] K. Forbes-Riley and D. Litman. Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication*, 53(9):1115–1136, 2011.

[11] K. Forbes-Riley and D. J. Litman. Adapting to student uncertainty improves tutoring dialogues. In *AIED*, pages 33–40, 2009.

[12] M. Glass, J. H. Kim, M. W. Evens, J. A. Michael, and A. A. Rovick. Novice vs. expert tutors: A comparison of style. In *MAICS-99, Proceedings of the Tenth Midwest AI and Cognitive Science Conference*, pages 43–49, 1999.

[13] A. C. Graesser and B. A. Olde. How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, pages 524–536, 2003.

[14] J. F. Grafsgaard, R. M. Fulton, K. E. Boyer, E. N. Wiebe, and J. C. Lester. Multimodal Analysis of the Implicit Affective Channel in Computer-Mediated Textual Communication. In *Proceedings of the 14th International Conference on Multimodal Interaction*, pages 145–152, 2012.

[15] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester. Automatically recognizing facial expression: Predicting engagement and frustration. In *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)*, pages 43–50, 2013.

[16] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester. Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014)*, pages 122–129, 2014.

[17] J. F. Grafsgaard, J. B. Wiggins, A. K. Vail, K. E. Boyer, E. N. Wiebe, and J. C. Lester. The additive value of multimodal features for predicting engagement, frustration, and learning during tutoring. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 42–49. ACM, 2014.

[18] E. Y. Ha, J. F. Grafsgaard, C. M. Mitchell, K. E. Boyer, and J. C. Lester. Combining Verbal and Nonverbal Features to Overcome the 'Information Gap' in Task-Oriented Dialogue. In *Proceedings of the 13th Annual SIGDIAL Meeting on Discourse and Dialogue*, pages 247–256, 2012.

[19] J. A. Harrigan and D. M. O'Connell. How do you look when feeling anxious? facial displays of anxiety. *Personality and Individual Differences*, 21(2):205–212, 1996.

[20] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, pages 139–183, 1988.

[21] M. E. Hoque, D. J. McDuff, and R. W. Picard. Exploring temporal patterns in classifying frustrated and delighted smiles. *Affective Computing, IEEE Transactions on*, 3(3):323–334, 2012.

[22] S. Kai, L. Paquette, R. S. Baker, N. Bosch, S. D'Mello, J. Ocumpaugh, V. Shute, and M. Ventura. A comparison of video-based and interaction-based affect detectors in physics playground.

[23] A. Kapoor and R. W. Picard. Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 677–682. ACM, 2005.

[24] D. Keltner. Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, 68(3):441–454, 1995.

[25] M. R. Lepper and M. Woolverton. The wisdom of practice: Lessons learned from the study of highly effective tutors. *Improving Academic Achievement: Impact of Psychological Factors on Education*, pages 135–158, 2002.

[26] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, M. Javier, and M. Bartlett. The computer expression recognition toolbox (CERT). In *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, pages 298–305, 2011.

[27] Z. Liu, V. Pataranutaporn, J. Ocumpaugh, and R. S. Baker. Sequences of frustration and confusion, and learning. In *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)*, pages 114–120, 2013.

[28] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin. Automatically detecting pain in video through facial action units. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 41(3):664–674, 2011.

[29] C. M. Mitchell, E. Y. Ha, K. E. Boyer, and J. C. Lester. Learner characteristics and dialogue: recognising effective and student-adaptive tutorial

strategies. *International Journal of Learning Technology 25*, 8(4):382–403, 2013.

[30] C. M. Mitchell, E. Y. Ha, K. E. Boyer, and J. C. Lester. Learner characteristics and dialogue: recognising effective and student-adaptive tutorial strategies. *International Journal of Learning Technology*, 8(4):382–403, 2013.

[31] M. Mortillaro, M. Mehu, and K. R. Scherer. Subtly Different Positive Emotions Can Be Distinguished by Their Facial Expressions. *Social Psychology and Personality Science*, 2(3):262–271, 2011.

[32] H. L. O'Brien and E. G. Toms. The development and evaluation of a survey to measure user engagement. 61(1):50–69, 2010.

[33] L. Paquette, B. Mott, K. Brawner, J. Rowe, J. Lester, R. Sottilare, R. Baker, J. DeFalco, and V. Georgoulas. Sensor-free or sensor-full: A comparison of data modalities in multi-channel affect detection. In *Under review for the 8th International Conference on Educational Data Mining,(under review)*.

[34] J. Piaget. *The origins of intelligence*. International University Press, 1952.

[35] C. P. Rosé, J. D. Moore, K. VanLehn, and D. Allbritton. A comparative evaluation of socratic versus didactic tutoring. *Proceedings of Cognitive Sciences Society*, pages 869–874, 2001.

[36] M. O. Z. San Pedro, E. L. Snow, R. S. Baker, D. S. McNamara, and N. T. Heffernan. Exploring dynamical assessments of affect, behavior, and cognition and math state test achievement.

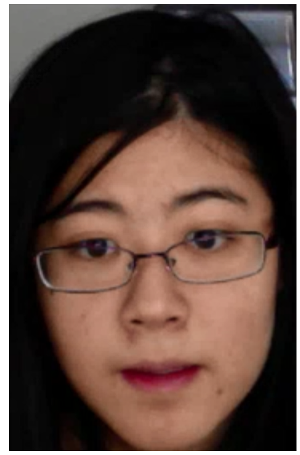
[37] A. K. Vail and K. E. Boyer. Adapting to Personality Over Time: Examining the Effectiveness of Dialogue Policy Progressions in Task-Oriented Interaction. In *Proceedings of the 15th Annual SIGDIAL Meeting on Discourse and Dialogue*, pages 41–50, 2014.

[38] A. K. Vail and K. E. Boyer. Identifying Effective Moves in Tutorial Dialogue: On the Refinement of Speech Act Annotation Schemes. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*, pages 199–209, 2014.

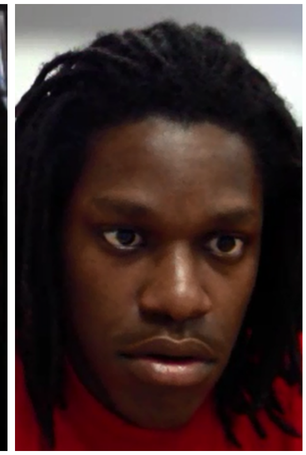
[39] A. K. Vail, J. F. Grafsgaard, J. B. Wiggins, J. C. Lester, and K. E. Boyer. Predicting learning and engagement in tutorial dialogue: A personality-based model. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 255–262. ACM, 2014.

[40] K. VanLehn. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221, 2011.

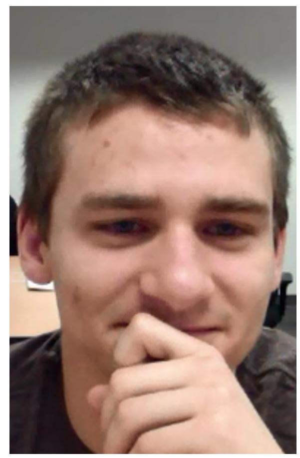
[41] M. Wixon, I. Arroyo, K. Muldner, W. Burleson, C. Lozano, and B. Woolf. The opportunities and limitations of scaling up sensor-free affect detection. pages 145–152, 2014.



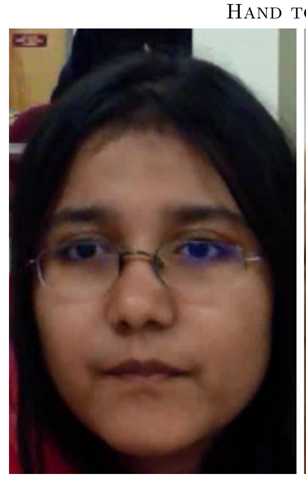
(a) AU2
OUTER BROW RAISER



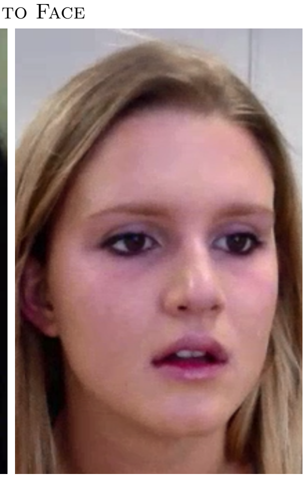
(b) AU5
UPPER LID RAISER



(c) AU10
UPPER LIP RAISER + ONE HAND TO FACE



(d) AU24
LIP PRESSOR



(e) AU26
JAW DROP

Figure 6: Sample frames from the student webcam illustrating the facial action unit features appearing in the predictive model for student engagement, as identified by FACET. Note that AU12 LIP CORNER PULLER (Figure 5b) also appears in these models.