Automatically Recognizing Facial Indicators of Frustration: A Learning-Centric Analysis

Joseph F. Grafsgaard, Joseph B. Wiggins, Kristy Elizabeth Boyer, Eric N. Wiebe*, and James C. Lester
Department of Computer Science
*Department of Science, Technology, Engineering, and Mathematics Education
North Carolina State University
Raleigh, NC, USA
{ jfgrafsg, jbwiggi3, keboyer, wiebe, lester } @ ncsu.edu

***Abstract*—Affective and cognitive processes form a rich substrate on which learning plays out. Affective states often influence progress on learning tasks, resulting in positive or negative cycles of affect that impact learning outcomes. Developing a detailed account of the occurrence and timing of cognitive-affective states during learning can inform the design of affective tutorial interventions. In order to advance understanding of learning-centered affect, this paper reports on a study to analyze a video corpus of computer-mediated human tutoring using an automated facial expression recognition tool that detects fine-grained facial movements. The results reveal three significant relationships between facial expression, frustration, and learning: 1) Action Unit 2 (outer brow raise) was negatively correlated with learning gain, 2) Action Unit 4 (brow lowering) was positively correlated with frustration, and 3) Action Unit 14 (mouth dimpling) was positively correlated with both frustration and learning gain. Additionally, early prediction models demonstrated that facial actions during the first five minutes were significantly predictive of frustration and learning at the end of the tutoring session. The results represent a step toward a deeper understanding of learning-centered affective states, which will form the foundation for data-driven design of affective tutoring systems.**

Keywords—affect; frustration; learning; computer-mediated tutoring; facial expression recognition; facial action units; intensity

I. Introduction

Cognitive and affective processes intertwine during learning, comprising a rich layer of emotional experience. Affective states often influence progress on learning tasks, resulting in positive or negative cycles of affect that impact learning outcomes [1–3]. Consequently, the influence of affective phenomena on learning has led to a recognized need to understand the occurrence, timing, and impact of cognitive-affective states during learning [1–5]. Developing a clear understanding of these phenomena is critical to informing the design of affective tutorial interventions [1–3].

Numerous studies have investigated learning-centered cognitive-affective states such as frustration, anxiety, boredom (or disengagement), confusion, delight, eureka (or a-ha moments), excitement, flow (or engaged concentration), and surprise [1–7]. Each of these states contributes to the complex intermingling of cognitive and affective processes inherent in learning. This paper focuses on frustration, which plays a central role in learning, possibly hindering it [1], [3]. When students are unable to surmount difficulties during learning tasks, they may remain in a "state of stuck" [1], [6], [7]. Similarly, if students are unable to reconcile confusion induced by new concepts, they may transition to frustration [2]. Thus, automatic detection or prediction of frustration is vital to designing affective tutorial interventions that alleviate frustration and foster learning.

Facial expression has proven particularly useful for investigating affect, in large part because of the ubiquity of facial expression in human experience and the non-invasiveness of video recording [2], [4], [5]. In many studies, the Facial Action Coding System (FACS), which enumerates possible movements of the human face, is used to manually annotate facial movements that comprise expressions of emotion [8], [9]. This approach has been used to study facial displays of learning-centered cognitive-affective states [2], [4], [5]. Often, facial expressions are recorded and then evaluated at moments of self-reports or judged affective events using FACS [2]. In this work, we aim to automatically detect facial action units that are related to frustration.

There have been many advances in automated facial expression recognition research in recent years [10], [11]. Widely used techniques have ranged from facial feature tracking to systems that automatically interpret emotions [10–12]. Facial feature tracking methods offer a convenient face mesh representation that can be used as input for constructing machine-learned models of facial expression [10], [11]. Systems that automatically interpret emotions are useful in many domains, but prior research has shown that the most frequently studied emotions (e.g., Ekman emotions [9]) are rarely present in learning [1], [2]. Prominent "off-the-shelf" facial expression tracking systems have tended to focus on recognizing Ekman emotions (e.g., happy, sad, angry) [13], [14]. More recently, there has been significant progress in automatically detecting FACS facial action units that may be correlated with learning-centered affective states [15–17]. One automated FACS coding system, the Computer Expression Recognition Toolbox (CERT), has recently been made available to the research community [15].

We applied CERT to a corpus of computer-mediated human-human tutoring in order to investigate relationships among automatically detected facial action units, affective outcomes, and learning gains. Based on frequencies of manually labeled facial action units in a previously analyzed computer-mediated tutoring corpus within the same domain [4], the five most frequently occurring action units were selected for automated analyses: AUs 1, 2, 4, 7, and 14. An

978-0-7695-5048-0/13 $26.00 © 2013 IEEE
DOI 10.1109/ACII.2013.33

initial analysis was conducted to provide a large-scale validation of CERT output versus manual FACS annotations and to produce exploratory predictive models of affective outcomes and pre-post learning gain, as reported in [18]. The present analyses built upon those results with a focused investigation through the lens of learning gain. First, student self-reported frustration was found to be negatively correlated with student learning. Second, correlations of facial expression, frustration, and learning produced three primary results: 1) AU2 (outer brow raise) was negatively correlated with learning gain, 2) AU4 (brow lowering) was positively correlated with frustration, and 3) AU14 (mouth dimpling) was positively correlated with both frustration and learning gain. These results represent possible indicators of affective states as they occur over time, such as frustration, anxiety, confusion, and effortful thought. Third, models were constructed to inform further study on early prediction of frustration and learning, demonstrating that facial actions within the first five minutes of a tutorial interaction are significantly predictive of student self-reports at the end of the session. These analyses suggest that current automated facial action unit tracking is sufficiently accurate to support wide-scale application in intelligent tutoring systems and education research. Further research in this vein will inform the next generation of affective tutorial interventions that respond to moment-by-moment frustration and other learning-centered cognitive-affective states.

II. Related Work

Studies of facial expressions related to learning-centered cognitive-affective states can be categorized into one of three paradigms: 1) observation and annotation of affective behaviors; 2) investigation of facial action units involved in learning-centered affect; and 3) application of automated methods to detect affective states. We review prior work in each of these categories, with a focus on facial expression recognition in the third category due to space constraints. However, we also note the substantial prior work on predicting frustration from physiology [6], [7].

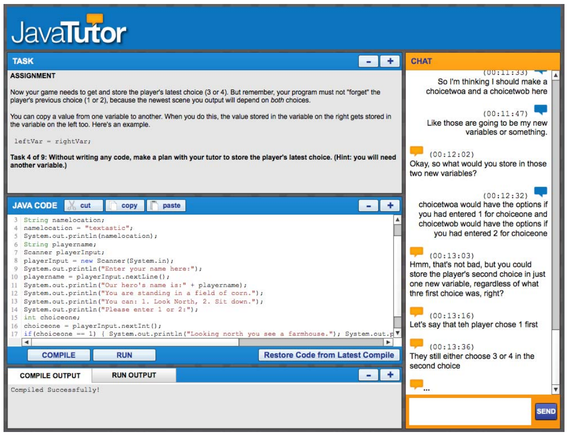
The first category of studies involves observing and annotating affective behavior, and often represents a precursor to further analyses of learning-centered affect. Prior to applying automated methods to detect student affective states during interactions with Wayang Outpost (a mathematics intelligent tutoring system), Woolf et al. observed student behaviors including head nodding/leaning, postural movement, verbalization, and smiling, which supported further study of arousal and valence [3]. Afzal & Robinson studied affect in a naturalistic video corpus taken during self-study of tutorial materials and a complex mental task, multiple coders were used to label emotions [19]. The coders identified confusion, happiness, interest, and surprise as the most frequent cognitive-affective states. Lastly, Baker et al. have used classroom observations to identify and analyze moment-by-moment affect during student interactions with cognitive tutors and other educational software [1]. The observation protocol has been developed over several years, and involves viewing students through peripheral vision to interpret their posture, facial expression, gesture, speech, and eye gaze. The protocol has been applied to student populations throughout the world, and has provided key insight into student affective states and learning, such as the detrimental nature of boredom. A key difference with the present study is that we aim to automatically detect learning-centered affect through facial action unit tracking, which is particularly applicable to computer-mediated tutoring.

The second category of studies involves investigating facial action units in learning-centered affect. These studies yield detailed data for designing affective tutoring systems. In a rich line of research, D'Mello and colleagues have compiled correlations of facial action units and self-reported and judged affect; for example, in a study of seven students' emote-alouds during interaction with AutoTutor (a natural language intelligent tutoring system that has been used in multiple domains), FACS coders annotated video at moments of students' emote-alouds [2]. In the same work, multiple judges annotated affect from videos of twenty-eight students' tutoring sessions with AutoTutor. The FACS labels of both studies were compared, identifying correlations of AU1 (inner brow raising) and AU2 (outer brow raising) with frustration, and correlations of AU4 (brow lowering) and AU7 (eyelid tightening) with confusion. In a previous study that we conducted, manual FACS coding was applied to a computer-mediated human-human tutoring corpus [4], [5]. In analysis of AU4 of fourteen students, a hidden Markov model (HMM) was machine-learned to investigate the role of confusion within the context of the tutorial dialogue and learning task [5]. The HMM accurately predicted AU4 events using the combination of prior AU4, dialogue, and task performance. In further work, we constructed a descriptive HMM from seven student sessions with manual FACS annotations of 16 AUs [4]. The analyses presented in this paper differ from prior studies in that we applied automated facial action unit tracking to identify correlations of facial expression, frustration, and learning. Manual FACS annotation is notably labor-intensive, so methods applying automated FACS coding can yield a substantially higher sample size across a broad set of facial expressions.

There have been few studies in the third category, which focuses on automatically detecting facial expressions of learning-centered affect. Woolf et al. tracked cognitive-affective states of students interacting with Wayang Outpost using the MindReader tracking software [20]. The MindReader system was trained on posed facial expressions and head movements of states such as interested or concentrating [12]. MindReader tracking of interest was found to improve predictive models of student self-reported confidence and excitement, while tracked interest did, in fact, improve predictive models of student self-reported interest [20]. In other work on automated detection, the authors of CERT have used it to track facial action units related to learning-centered affective states [21], [22]. For instance, CERT was used to track facial expressions of students interacting with a human tutor operating an iPad interface during cognitive game tasks (a Wizard-of-Oz design) [22]. Additionally, CERT has been used to track children's facial expressions during a cognitive task [21]. In both cases, CERT output was used as a relative comparison measure (i.e., the amount and type of facial movement before, during, and after performing a task). While this provides insight into facial expressions at meaningful moments, the cognitive tasks were simplified and may not have captured the full complexity of cognitive and affective processes involved during learning in an academic scenario. The present work represents a novel investigation of automatically detected facial action units involved in learning-centered affect during tutoring, with indications of specific facial action units correlated with frustration and learning outcomes.

III. Computer-Mediated Human Tutoring Corpus

The corpus consists of computer-mediated tutorial dialogue for computational concepts. Students (*N*=67) and tutors interacted through a web-based interface that provided learning tasks, an interface for computer programming, and textual dialogue. The participants were university students in the United States, with average age of 18.5 years (*stdev*=1.5). The students voluntarily participated for course credit in an introductory engineering course, but no prior computer science knowledge was assumed or required. Substantial self-reported prior programming experience was an exclusion criterion. Each student was paired with a tutor for a total of six sessions on different days, each session limited to forty minutes. Recordings of the sessions included database logs, webcam video, skin conductance, and Kinect depth video. The JavaTutor interface is shown in Fig. 1 and the recording setup is shown in Fig. 2.



ASSIGNMENT

Now your game needs to get and store the player's latest choice (3 or 4). But remember, your program must not "forget" the player's previous choice (1 or 2), because the newest scene you output will depend on *both* choices.

You can copy a value from one variable to another. When you do this, the value stored in the variable on the right gets stored in the variable on the left too. Here's an example.

```
leftVar = rightVar;
```

Task 4 of 9: Without writing any code, make a plan with your tutor to store the player's latest choice. (Hint: you will need another variable.)

```
String namelocation;
namelocation = "textastic";
System.out.println(namelocation);
String playername;
Scanner playerInput;
playerInput = new Scanner(System.in);
System.out.println("Enter your name here:");
playername = playerInput.nextLine();
System.out.println("Our hero's name is:" + playername);
System.out.println("You are standing in a field of corn.");
System.out.println("You can: 1. Look North, 2. Sit down.");
System.out.println("Please enter 1 or 2:");
int choiceone;
choiceone = playerInput.nextInt();
if(choiceone == 1) { System.out.println("Looking north you see a farmhouse."); System.out.p
```

```
Compiled Successfully!
```

(00:11:33)
So I'm thinking I should make a choicetwoa and a choicetwob here

(00:11:47)
Like those are going to be my new variables or something.

(00:12:02)
Okay, so what would you store in those two new variables?

(00:12:32)
choicetwoa would have the options if you had entered 1 for choiceone and choicetwob would have the options if you had entered 2 for choiceone

(00:13:03)
Hmm, that's not bad, but you could store the player's second choice in just one new variable, regardless of what thre first choice was, right?

(00:13:16)
Let's say that teh player chose 1 first

(00:13:36)
They still either choose 3 or 4 in the second choice

...

Figure 1. The JavaTutor interface

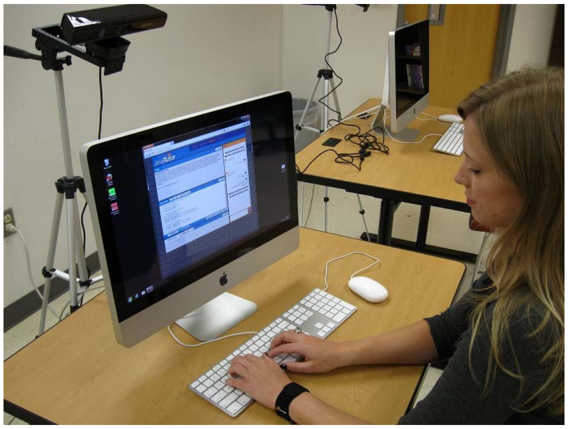


Figure 2. Student workstation with Kinect depth camera, skin conductance bracelet, and computer with webcam

In this study, we analyzed the webcam video corpus of the first lesson for each student. The study coordinators started the recordings at the beginning of each tutoring session. Thus, the students were aware of the recordings. However, once started, the recording windows were automatically hidden, so the students did not see themselves during the tutoring sessions. Additionally, when reviewing the recordings in the process of our analyses, we observed that students did not attend to the recording devices (webcam and Kinect), which indicates that the recordings were unobtrusive. The tutoring video corpus is comprised of approximately four million video frames totaling thirty-seven hours across the first tutoring lesson. Two session recordings were missing due to human error (*N*=65). The recordings were taken at 640*x*480 pixel resolution and thirty frames per second.

Before each session, students completed a content-based pretest. After each session, students answered a post-session survey and posttest (identical to the pretest). The post-session survey items were designed to measure several aspects of engagement and cognitive load. The survey consisted of a User Engagement Survey (UES) [23] with Focused Attention, Endurability, and Involvement subscales, and the NASA-TLX workload scale [24], which consisted of response items for Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level. The UES subscales were each comprised of multiple Likert-style items, while each NASA-TLX item was self-reported on a scale from zero to one hundred. As noted by one of the authors of NASA-TLX in a retrospective survey of studies using the scale, individual response items can be used separately to identify specific dimensions of task workload [25]. Thus, our analysis considers the response items, including the item for Frustration Level, independently from the other survey items.

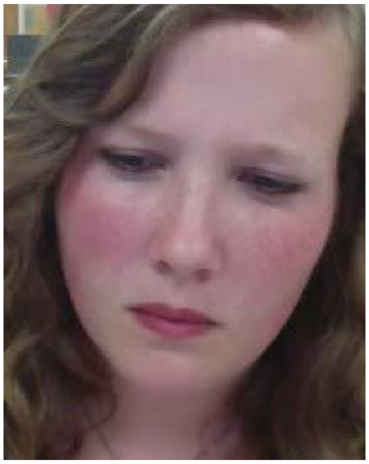
IV. Facial Expression Recognition

The Computer Expression Recognition Toolbox (CERT) [15] was used in this study because it allows frame-by-frame tracking of a wide variety of facial action units. CERT finds faces in a video frame, locates facial features for the nearest face, and outputs weights for each tracked facial action unit using support vector machines [16]. CERT has been validated for use with both adults and children [15], [16], [21].

Based on observations from prior studies in task-oriented tutoring, we selected a subset of the 30 facial action units that CERT detects as the focus of the present analyses. The subset of facial action units was informed by a prior naturalistic tutoring video corpus that was manually annotated by certified FACS coders [4]. The five most frequently occurring facial action units were selected for the present study. These were AUs 1, 2, 4, 7, and 14. We observed that raw CERT output varied significantly across individuals, as noted by the authors of CERT [16], [22]. Thus, we adjusted the output values by subtracting the baseline (average) output value by facial action unit for each student.

The naturalistic tutoring video corpus described above [4] was processed using CERT, enabling comparison with the manual annotations of AU presence and absence. In order to interpret CERT output as indicating presence or absence of a facial action unit, a detection threshold of 0.25 was empirically determined. For instance, the average adjusted CERT output value for AU7 present in the prior corpus was 0.29, while the average for AU7 absent was -0.01. In comparison, the average raw CERT output for AU7 present was 0.25, while the average for AU7 absent was 0.19. In our initial large-scale validation of CERT output with manual annotations in the prior corpus, we found that baseline-adjusted CERT output produced more consistent values in the presence or absence of the five selected AUs compared to raw CERT output [18]. Thus, a combination of baseline adjustment of CERT output and an empirically-determined detection threshold allows for comparison of facial action units across individuals.

Fig. 3 provides detailed comparison of raw CERT output versus baseline-adjusted CERT output of AU7. Note that the raw CERT output indicates that AU14 (mouth dimpling) is present, though it is apparent to a certified FACS coder that the action unit is not present in the image. The adjusted CERT output correctly indicates that AU7 is present and AU14 is absent (based on the 0.25 threshold). Fig. 4 shows adjusted CERT output from an example of AU 2, at three moments of the facial expression event: just before onset, apex (most intense video frame), and just after offset. Fig. 5 shows adjusted CERT output for AUs 1, 4, and 14.



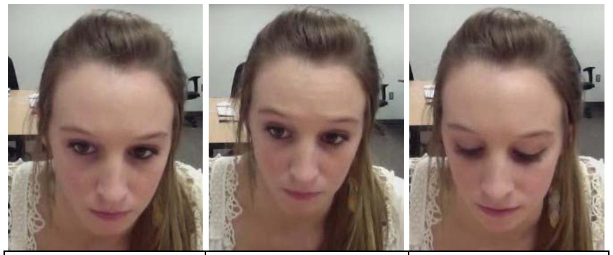
RAW CERT OUTPUT

Action Unit	Value
AU1	-1.36
AU2	-0.67
AU4	0.05
AU7	**0.52**
AU14	**1.39**

ADJUSTED CERT OUTPUT

Action Unit	Value
AU1	-0.22
AU2	-0.27
AU4	0.11
AU7	**0.26**
AU14	-0.04

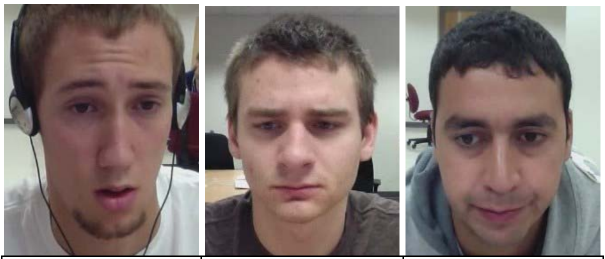
Figure 3. A comparison of AU7 with raw and baseline-adjusted CERT output



AU2 Onset: Outer brow raiser	**AU2 Apex:** Outer brow raiser	**AU2 Offset:** Outer brow raiser

AU1(-0.76) AU2(-0.21)
AU4(-0.09) AU7(-0.24)
AU14(0.13)
	AU1(0.17) **AU2(0.27)**
	AU4(0.08) AU7(-0.09)
	AU14(-0.53)
		AU1(-0.94) AU2(-0.21)
		AU4(0.12) AU7(0.09)
		AU14(0.00)

Figure 4. The onset, apex and offset of an AU2 event
with baseline-adjusted CERT output



AU1 and AU4 Apex: Inner brow raiser and brow lowerer	**AU4 Apex:** Brow lowerer	**AU14 Apex:** Dimpler

AU1(0.80) AU2(-0.12)
AU4(0.25) AU7(-0.23)
AU14(-0.06)
	AU1(-0.02) AU2(0.07)
	AU4(0.47) AU7(0.08)
	AU14(-0.85)
		AU1(-0.02) AU2(0.00)
		AU4(-0.04) AU7(0.00)
		AU14(0.46)

Figure 5. Examples of AU1+AU4, AU4, and AU14
with baseline-adjusted CERT output

V. FRUSTRATION, LEARNING AND FACIAL EXPRESSION

We used correlational analyses to compare affective post-session surveys, learning gains, and facial expression. First, we identified that frustration was the only post-session affect self-report that correlated with learning gain. Second, we focused on how facial expression throughout the sessions correlated with frustration and learning gain. Finally, we examined facial expressions at the beginning and end of the sessions and their statistical relationships with frustration and learning gain.

Normalized learning gains were calculated as follows, if posttest score was greater than pretest score:

$$NLG = \frac{Posttest - Pretest}{1 - Pretest}$$

Otherwise, the following formula was used:

$$NLG = \frac{Posttest - Pretest}{Pretest}$$

Thus, negative learning gains were possible, although 61 out of the 65 students had positive learning gain (*min*=-0.29, *max*=1.00). Correlational analyses of the post-session affective survey scales and learning gains were conducted to identify potential relationships between affect and learning. Frustration was the only affect self-report to correlate with learning gain, with higher frustration corresponding to lower learning gain (TABLE I).

TABLE I. CORRELATIONS OF AFFECTIVE POST-SURVEY SCALES AND NORMALIZED LEARNING GAIN

Affect Survey Scale	*r*	*p*
ENGAGEMENT	0.14	0.288
MENTAL DEMAND	0.02	0.871
PHYSICAL DEMAND	-0.06	0.610
TEMPORAL DEMAND	0.03	0.830
PERFORMANCE	0.18	0.160
EFFORT	0.04	0.732
FRUSTRATION LEVEL	**-0.30**	**0.015**

Because of the important relationship between frustration and learning, we focused our present analyses of facial expression on frustration and learning gain. The analyses of facial expression consider two features for each facial action unit: average intensity (average magnitude of CERT output values that were above the detection threshold) and relative frequency (percent of tracked frames that were above the detection threshold) of each facial action unit. These features were calculated across each tutoring session, resulting in ten feature values per student. We first considered correlations across entire tutoring sessions, and then examined correlations for the first five minutes and last five minutes of the sessions. The beginning of a session may inform early prediction, while the end may more closely reflect self-reports due to temporal proximity. We applied a statistical correction for multiple tests, arriving at a Bonferroni *p*-value threshold of 0.0025 for each set of analyses of facial movements: entire session, beginning of session, and end of session. This more stringent threshold controls for the risk of false positives, and is intended to increase the generalizability of the findings. Results that were significant to this threshold are displayed in bold, and all other correlations with $p<0.05$ are shown.

The initial analyses correlated facial action units throughout the tutoring sessions with normalized learning gains and post-session self-reports of frustration (TABLE II). Intensity of AU4 was positively correlated with frustration. Thus, greater intensity of AU4 corresponded with higher self-report of frustration. Both intensity and frequency of AU2 were negatively correlated with normalized learning gain, with only AU2 intensity significant after Bonferroni correction.

TABLE II. CORRELATIONS OF FRUSTRATION, LEARNING AND FACIAL ACTION UNITS THROUGHOUT THE TUTORING SESSION

Action Unit Variable	Tutoring Outcome	*r*	*p*
AU 4 Avg. Intensity	Frustration	0.31	0.011
AU 2 Avg. Intensity	**Norm. Learn. Gain**	**-0.38**	**0.002**
AU 2 Relative Freq.	Norm. Learn. Gain	-0.27	0.029

Analyses were conducted to examine whether facial expressions in the first five minutes of tutorial interaction were correlated with frustration and normalized learning gain. Four results emerged (TABLE III): intensity of AU14 was positively correlated with frustration and frequency of AU2 was negatively correlated with normalized learning gain. Additionally, intensity of AU4 was positively correlated with frustration, and intensity of AU2 was negatively correlated with normalized learning gain. The correlations involving AU2 intensity and AU4 intensity retained their statistical significance after Bonferroni correction.

TABLE III. CORRELATIONS OF FRUSTRATION, LEARNING AND FACIAL ACTION UNITS IN THE FIRST FIVE MINUTES OF TUTORING

Action Unit Variable	Tutoring Outcome	*r*	*p*
AU 4 Avg. Intensity	**Frustration**	**0.41**	**0.001**
AU 14 Avg. Intensity	Frustration	0.32	0.010
AU 2 Avg. Intensity	**Norm. Learn. Gain**	**-0.38**	**0.002**
AU 2 Relative Freq.	Norm. Learn. Gain	-0.28	0.023

There was one significant result in the correlational analysis of frustration, learning, and facial action units in the last five minutes of tutoring. Relative frequency of AU14 was found to positively correlate with normalized learning gain (r=0.52, p<0.001). This result was significant after Bonferroni correction.

VI. EARLY PREDICTION OF FRUSTRATION AND LEARNING

Linear regression models were constructed to predict frustration and normalized learning gain from facial expressions in the first five minutes of tutoring. These models are intended to inform the use of facial expression features for early prediction. Further development of features would be necessary to drive affective tutorial intervention.

Both models were constructed using the significantly correlated facial action unit features from Section V. The early prediction model of frustration is shown in TABLE IV. The model R^2 value corresponds to r=0.49, and the model effect is greater than either feature alone. The root mean squared error (RMSE) value indicates the overall magnitude of error. The RMSE of this model shows that it would not distinguish well between similar frustration levels, but would perform well at distinguishing between very high or low levels of frustration.

The model for early prediction of normalized learning gain is shown in TABLE V. The model R^2 value corresponds to r=0.40. Thus, the model effect is similar to the most explanatory feature, AU2 intensity. The significance values for the features also show that AU2 frequency did not significantly explain variance beyond AU2 intensity. The RMSE of this model is similar to that of the early prediction model of frustration. Very high or low values may be accurately distinguished, but it is likely to misidentify similar values.

TABLE IV. EARLY PREDICTION MODEL OF FRUSTRATION

Frustration Level =	***p***
81.72 * AU4 Intensity	.002
38.62 * AU14 Intensity	.022
Intercept = -41.69	.002
RMSE = 21% of range in self-reports **Model R^2** = 0.24	

TABLE V. EARLY PREDICTION MODEL OF NORM. LEARNING GAIN

Normalized Learning Gain =	***p***
-1.45 * AU2 Intensity	0.020
-0.46 * AU2 Relative Frequency	0.273
Intercept = 1.12	< 0.0001
RMSE = 24% of range in outcomes **Model R^2** = 0.16	

VII. DISCUSSION

We have presented results that demonstrate important relationships among frustration, learning and facial expression within a corpus of computer-mediated human-human tutoring. In this section we focus on interpreting these findings. Two notable characteristics of the corpus facilitate this interpretation. First, the corpus reflects few social effects on nonverbal behavior due to remote dialogue because the students and tutors did not see one another. Nonverbal behaviors that are common in face-to-face communication, such as *emblems* (e.g., thumbs-up gesture), *illustrators* (e.g., gesticulating to illustrate an idea during speech), and *regulators* (e.g., gesturing for a conversational participant to speak) were not displayed [26]. Second, the video recording of students was accomplished discreetly (e.g., not making noise or displaying a red light during recording). If the act of recording were obtrusive, students would likely become distracted and self-conscious, perhaps resulting in inhibition of facial expression. However, as noted in SECTION III, we observed that students did not attend to the recording devices during the tutoring sessions.

The analyses highlight ways in which intensity and frequency of facial action unit displays are associated with normalized learning gains and summative self-reports of frustration. Each facial action unit has been explored in prior research. Thus, we consider the key results in light of past findings and theoretical implications of each facial action unit.

Both intensity and frequency of outer brow raising (AU2) were negatively correlated with normalized learning gain, based on AU2 displays at the beginning of tutoring sessions and throughout tutoring sessions. Based on prior literature, AU2 may be associated with frustration, surprise, or anxiety. AU2 has been identified as a component of frustration (along with AU1) in prior intelligent tutoring systems literature, with correlations of AU1 and AU2 displays and students' emote-aloud self-reports of frustration [2]. Similarly to frustration, AU1 and AU2 together are components of the prototypical expression of surprise [9]. However, frustration and surprise do not seem consistent with the results of the correlational analyses because AU1 was absent from the significant correlations.

Anxiety has been linked to prototypical displays of fear (AU1+AU2+AU4+AU5+AU25; AU5 is eyelid opening, AU25 is mouth opening) [27]. While this combination of AUs seems similar to those of frustration and surprise, the presence of AU4 introduces a conflicting movement of the brow that may impact detection of the expression. Fig. 4 shows an example of AU1+AU2+AU4 (also shown are the moments before and after the facial expression—just before onset and after offset). The CERT output values from the apex of the facial expression show an interaction between AU1 and AU4. AU1 raises the inner eyebrows, while AU4 lowers the inner brow. The result of AU1+AU2+AU4 is tensing of the inner brow with creasing across the forehead, as is apparent in Fig. 4 to a FACS coder. This conflict of facial movements at the inner brow may result in reduced CERT output values for both AU1 and AU4. This complication of CERT output may explain how only AU2 was negatively correlated with normalized learning gain. It also indicates that anxiety may be the most consistent interpretation of AU2.

Brow lowering (AU4) intensity at the beginning of sessions and throughout sessions was positively correlated with summative self-reports of frustration. AU4 has long been noted as an indicator of mental effort, notably mentioned by Darwin [21]. AU4 has also been correlated with self-reports and judgments of confusion in intelligent tutoring systems research [2]. In this interpretation, AU4 at the beginning of sessions may have indicated effortful thinking and confusion. It is possible that such confusion may have gone unresolved, resulting in frustration.

Mouth dimpling (AU14) intensity at the beginning of sessions positively correlated with frustration, while AU14 frequency at the end of sessions positively correlated with normalized learning gain. Unilateral AU14 is a component of a prototypical expression of contempt [28]. However, students were observed to frequently display bilateral AU14 in our corpus, as in Fig. 5. Prior literature provides slight evidence of correlation between AU14 and frustration, with a statistical trend that AU14 occurred during student emote-aloud self-reports of frustration [2]. In the same study, expert judges did not identify AU14 as an indicator of frustration. We observed that AU14 appeared as a frequent 'mouth fidgeting' movement in our corpus. Thus, AU14 may be easily overlooked by judges of emotion as noise, since it occurs frequently over time, similar to blinking or brow lowering. As an affective feature, AU14 does seem to be a facial indicator that repeatedly occurs over time and coincides with a thoughtful state, as suggested by a prior study [21]. The correlation between intensity of AU14 displays at the beginning of sessions and frustration may parallel the discussion of AU4 above. It is possible that effortful thought or confusion transitioned to frustration [1], [2]. Additionally, the correlation of AU14 at the end of the session with normalized learning gain suggests that students who were concentrating more at the end of the session tended to have increased learning gain.

The early prediction results illustrate that facial expression at the beginning of tutoring sessions may provide a useful set of features for early diagnosis of affective states related to post-session outcomes. The models presented here are descriptive and are not designed to be used for intervention, but richer models may be constructed using machine learning techniques. Further models may incorporate timing of facial expression and learning task context to increase predictive accuracy.

VIII. Conclusion and Future Work

Cognition and emotion pervade human experience, with tutorial interactions being particularly rich in cognitive-affective phenomena. The close coupling of cognitive and affective dimensions require that integrated models of cognition and affect inform the design of intelligent tutoring systems. In a step toward realizing this goal, we have investigated relations among automatically detected facial action units, affective outcomes, and learning gains in a computer-mediated human tutoring corpus. This investigation first identified several strong statistical relationships between AU2 (outer brow raise) with learning, AU4 (brow lowering) with frustration, and AU14 (mouth dimpling) with both frustration and learning. Additionally, observing these facial action units during the first five minutes of tutoring contributes to predicting self-reports at the end of the session. It is hoped that similar analyses may continue the trend toward developing a greater understanding of learning-centered affective states to lay the groundwork for data-driven design of affective tutorial intervention.

This study highlights the potential for large-scale analyses of moment-by-moment affect in tutoring. Aside from frequency and intensity, there remains much to discover about facial expressions of learning-centered affect. For example, the temporal characteristics of facial expression in tutoring are largely unexplored. We have noted that AU4 (brow lowering) and other thoughtful expressions may extend over long periods of time. Automated facial action unit tracking allows for close examination of persistence of cognitive-affective states, so theoretical transitions among learning-centered affective states may be tested at fine temporal granularity. Additionally, coincidence of learning task context and facial expression may also be investigated. With a solid, statistically grounded foundation for learning-centered affect, future intelligent tutoring systems may achieve dynamic affective responsiveness that rivals that of the best human tutors.

Acknowledgments

This work is supported in part by the North Carolina State University Department of Computer Science and the National Science Foundation through Grant DRL-1007962 and the STARS Alliance Grant CNS-1042468. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the authors, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

References

[1] R. S. J. d. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser, "Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments," *International Journal of Human-Computer Studies*, vol. 68, no. 4, pp. 223–241, Apr. 2010.

[2] S. K. D'Mello, S. D. Craig, and A. C. Graesser, "Multi-Method Assessment of Affective Experience and Expression during Deep Learning," *International Journal of Learning Technology*, vol. 4, no. 3/4, pp. 165–187, 2009.

[3] B. P. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. G. Cooper, and R. W. Picard, "Affect-Aware Tutors: Recognising and Responding to Student Affect," *International Journal of Learning Technology*, vol. 4, no. 3–4, pp. 129–164, 2009.

[4] J. F. Grafsgaard, K. E. Boyer, and J. C. Lester, "Toward a Machine Learning Framework for Understanding Affective Tutorial Interaction," in *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, 2012, pp. 52–58.

[5] J. F. Grafsgaard, K. E. Boyer, and J. C. Lester, "Predicting Facial Indicators of Confusion with Hidden Markov Models," in *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction*, 2011, pp. 97–106.

[6] S. W. McQuiggan, S. Lee, and J. C. Lester, "Early Prediction of Student Frustration," in *Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction*, 2007, pp. 698–709.

[7] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic Prediction of Frustration," *International Journal of Human-Computer Studies*, vol. 65, no. 8, pp. 724–736, Aug. 2007.

[8] P. Ekman and W. V. Friesen, *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press, 1978.

[9] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System: Investigator's Guide*. Salt Lake City, USA: A Human Face, 2002.

[10] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[11] R. A. Calvo and S. K. D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010.

[12] R. El Kaliouby and P. Robinson, "The Emotional Hearing Aid: an Assistive Tool for Children with Asperger Syndrome," *Universal Access in the Information Society*, vol. 4, no. 2, pp. 121–134, Aug. 2005.

[13] T. Ruf, A. Ernst, and C. Kublbeck, "Face Detection with the Sophisticated High-speed Object Recognition Engine (SHORE)," in *Microelectronic Systems*, 2011, pp. 243–252.

[14] M. J. den Uyl and H. van Kuilenburg, "The FaceReader: Online Facial Expression Recognition," in *Proceedings of Measuring Behavior 2005*, 2008, pp. 589–590.

[15] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The Computer Expression Recognition Toolbox (CERT)," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2011, pp. 298–305.

[16] T. Wu, N. J. Butko, P. Ruvolo, J. Whitehill, M. S. Bartlett, and J. R. Movellan, "Multi-Layer Architectures for Facial Action Unit Recognition," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 42, no. 4, pp. 1027–1038, 2012.

[17] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated Facial Action Coding System for Dynamic Analysis of Facial Expression in Neuropsychiatric Disorders," *Journal of Neuroscience Methods*, vol. 200, no. 2, pp. 224–238, 2011.

[18] 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*, 2013.

[19] S. Afzal and P. Robinson, "Natural Affect Data - Collection & Annotation in a Learning Context," in *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction*, 2009, pp. 1–7.

[20] I. Arroyo, D. G. Cooper, W. Burleson, B. P. Woolf, K. Muldner, and R. M. Christopherson, "Emotion Sensors Go To School," in *14th International Conference on Artificial Intelligence in Education*, 2009, pp. 17–24.

[21] G. Littlewort, M. S. Bartlett, L. P. Salamanca, and J. Reilly, "Automated Measurement of Children's Facial Expressions during Problem Solving Tasks," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2011, pp. 30–35.

[22] J. Whitehill, Z. Serpell, A. Foster, Y.-C. Lin, B. Pearson, M. Bartlett, and J. Movellan, "Towards an Optimal Affect-Sensitive Instructional System of Cognitive Skills," in *Proceedings of the Computer Vision and Pattern Recognition Workshop on Human Communicative Behavior*, 2011, pp. 20–25.

[23] H. L. O'Brien and E. G. Toms, "The Development and Evaluation of a Survey to Measure User Engagement," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 1, pp. 50–69, 2010.

[24] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. Amsterdam: Elsevier Science, 1988, pp. 139–183.

[25] S. G. Hart, "NASA-Task Load Index (NASA-TLX); 20 Years Later," in *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 2006, vol. 50, no. 9, pp. 904–908.

[26] M. Pantic, A. Pentland, A. Nijholt, and T. Huang, "Human Computing and Machine Understanding of Human Behavior: A Survey," in *Proceedings of the 8th International Conference on Multimodal Interaction*, 2006, pp. 239–248.

[27] J. A. Harrigan and D. M. O'Connell, "How Do You Look When Feeling Anxious? Facial Displays of Anxiety," *Personality and Individual Differences*, vol. 21, no. 2, pp. 205–212, 1996.

[28] D. Matsumoto and P. Ekman, "The Relationship Among Expressions, Labels, and Descriptions of Contempt," *Journal of Personality and Social Psychology*, vol. 87, no. 4, pp. 529–540, Oct. 2004.