

Gender Differences in Facial Expressions of Affect During Learning

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

Affective support is crucial during learning, with recent evidence suggesting it is particularly important for female students. Facial expression is a rich channel for affect detection, but a key open question is how facial displays of affect differ by gender during learning. This paper presents an analysis suggesting that facial expressions for women and men differ systematically during learning. Using facial video automatically tagged with facial action units, we find that despite no differences between genders in incoming knowledge, self-efficacy, or personality profile, women displayed one lower facial action unit significantly more than men, while men displayed brow lowering and lip fidgeting more than women. However, numerous facial actions including brow raising and nose wrinkling were strongly correlated with learning in women, whereas only one facial action unit, eyelid raiser, was associated with learning for men. These results suggest that the entire affect adaptation pipeline, from detection to response, may benefit from gender-specific models in order to support students more effectively.

Keywords

Affect, facial expression, gender effects, learning, intelligent tutoring systems.

1. INTRODUCTION

Modeling and adapting to users during learning has long been a central goal of the user modeling community [15, 38,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UMAP '16, July 13-17, 2016, Halifax, NS, Canada

© 2016 ACM. ISBN 978-1-4503-4370-1/16/07...\$15.00

DOI: <http://dx.doi.org/10.1145/2930238.2930257>

42]. These models of users frequently focus on cognitive aspects of learning, by modeling domain knowledge with techniques such as constraint-based modeling [38], knowledge tracing [15], and strategy learning for tutoring [9, 27].

Building on a rich history of modeling the cognitive aspects of learning, there is increasing awareness that the affective aspects of learning are perhaps equally important (e.g., [12, 39]). For example, adapting to students' uncertainty [9] and confusion [14] have been shown to significantly improve the effectiveness of tutoring. Recent years have thus seen increasing attention to assessing students' affective states while interacting with learning environments, including through self-reports [43] and through affect detectors that utilize log data [10].

Of all of these mechanisms for sensing or detecting affect, facial expressions are among the most useful channels because of their rich expressiveness and relative ease of automatic collection and analysis [13, 7]. Facial expressions have long been studied in psychology, where the findings have established the importance of the face in expressing emotion [20]. In contrast to the "basic emotions" that have been identified as central in everyday life—such as anger, fear, and happiness—there is growing consensus that, during learning, a subset of these emotions, along with learning-centric emotions such as boredom and confusion are highly prominent and influential [17, 4]. Empirical results have found that important facial expression differences emerge between students who find a learning experience frustrating versus those who find it engaging [25], and preliminary findings point to different patterns of facial expression depending on learners' age [24].

Complementing research on affect in learning, a host of studies have established that females and males express emotion differently on the face, and also exhibit differing levels of capability to recognize facial expressions of emotion. For example, women respond with more pronounced facial expression when viewing the facial expressions of others [16], and women can more accurately identify multiple emotions from facial expression even in the presence of minimal infor-

mation to identify an emotion [30]. In light of these results, a central open question emerges for user modeling during learning: *In what ways do females and males differ in their facial expression of emotion during learning?*

This paper investigates that question in the context of one-on-one tutoring for introductory computer programming. These results suggest that women and men display systematically different facial expressions during learning, and that modeling these facial expressions may lead to deeper understanding of affective learning processes.

2. RELATED WORK

2.1 Gender and Facial Expression

Research in psychology, biology, and physiology has long examined facial expressions and uncovered differences based on gender. Women are more facially expressive than men; for example, when viewing another person's facial expression, women show more pronounced facial reactions [16]. Men and women both show differences in emotion expression based on context: for example, while a smile is generally a genuine indicator of positive emotions, it may sometimes be contrived to meet social expectations. Gender plays a role in this difference, particularly when there is a power differential between two conversational partners [33]. In addition to differences in expression, there may also be important differences in emotion recognition between females and males. For example, when asked to identify multiple emotions from the facial expression of another person, women were significantly better at emotion identification than men [30]. Women also seem to be substantially better at decoding verbal cues than men [29]. Given these many differences, there is reason to believe that when it comes to facial expression, women express affect during learning differently than men and that gender differences may be important to consider in models of affect.

2.2 Facial Expression in Learning

Facial expression is one of the foremost channels with which to infer learning-centric emotion [18]. In a longstanding line of work involving Wayang Outpost, a mathematics tutoring system, researchers have categorized facial expression according to the affective state it suggests [44] and have attempted to predict interest and confidence using multimodal features including facial expression [2]. In studies with AutoTutor, a tutoring system with a conversational dialogue agent, students' emotions as evidenced by "emoting aloud" were modeled using facial expressions [11] and affect-adaptive versions of the system have been built that respond to emotions such as boredom, frustration, and confusion [19]. Facial expression as tutoring unfolds has also been shown to predict frustration and engagement that students self-report at the end of the tutoring session [26]. It is clear that analyses of facial expression can provide insight into affective processes that significantly influence learning.

2.3 Gender and Learning From Tutoring

Tutoring, both human-human tutoring and tutoring with intelligent systems, have been shown to be very effective means of support learning [41]. Recent results have suggested that women and men benefit from different approaches to tutoring. When different versions of an intelligent tutoring system for computer science were compared, women

reported significantly more engagement and less frustration with the version that offered affective support than with other versions of the system [40]. In another comparison of different intelligent tutoring systems for mathematics, women benefited from having choice within an intelligent tutoring system while men did not show a significant difference [27]. Previous studies have also demonstrated that embodied agents [1] and motivational scaffolding [3] are particularly beneficial for female students, and that females are more sensitive to the coordination between an agent's verbal and nonverbal behaviors [6]. Because women and men appear to benefit from different tutoring strategies, and because models of affect are highly informative for tutorial strategy choice, it is important for the field to devise affective models for learning that take gender into account.

3. HUMAN-HUMAN TUTORING STUDY

The data examined in the present analysis consists of a set of computer-mediated human-human interactions through a tutorial interface during a set of lessons in introductory computer science [28, 37]. The tutorial interface, displayed below in Figure 1, consists of four panes: the task description, the students source code, the compilation and execution output from the student's program, and the textual dialogue messages exchanged between the tutor and the student. Tutors' interactions with the student were limited to the exchange of textual messages and progressing between tasks. For an example of the dialogue exchanged between the tutor and student, see Table 1.

3.1 Participants

Student users ($N = 67$) were university students in the United States recruited from an introductory engineering course. There were 24 female students and 43 male students, with an average age of 18.5 years ($s = 1.5$ years). Most (58) of the students were college freshmen. Nearly all (59) of the students were majors in an engineering discipline, with 10 from computer science, 6 each from aerospace engineering and electrical and computer engineering, 5 each from mechanical and electrical, and the remaining students from disciplines such as civil, industrial, and nuclear engineering, as well as mathematics or physics. The human tutors ($N = 5$) were primarily graduate students with previous experience in tutoring introductory computer science.

3.2 Learning Outcomes

The students completed a series of six 40-minute lessons over the course of four weeks. The present analysis focuses on data from the first lesson. Immediately prior to and immediately following the completion of each session, students were administered a content-based pretest and an identical posttest. The average raw learning gain for the first lesson was 22.15% ($s = 17.7\%$); for female students the average was 23.65% ($s = 15.13\%$) and for males 21.29% ($s = 19.14\%$). Normalized learning gain was calculated according to Equation 1.

$$norm_gain = \begin{cases} \frac{post - pre}{1 - pre} & post > pre \\ \frac{post - pre}{pre} & post \leq pre \end{cases} \quad (1)$$

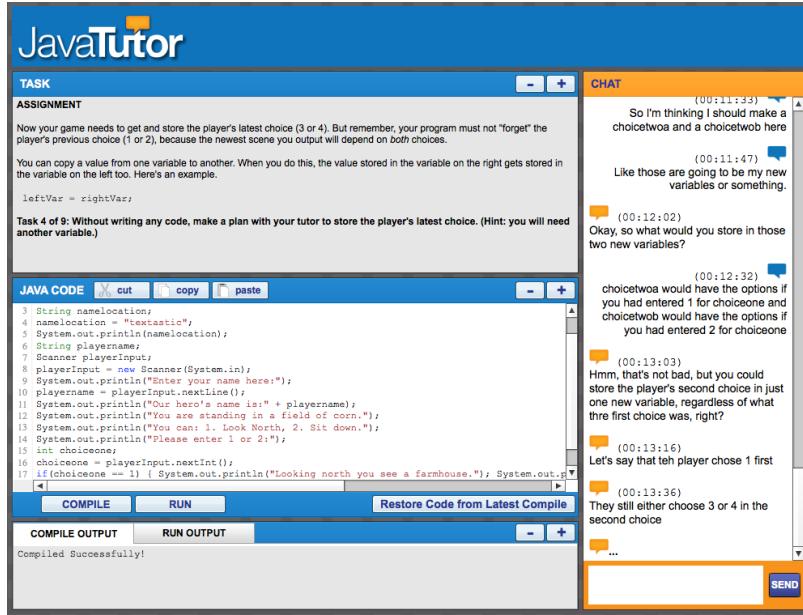


Figure 1: The tutorial interface for introductory Java programming.

3.3 Facial Expression Data Collection

Interaction data was collected through a set of multi-modal sensors, including a Kinect depth camera, an integrated webcam, and a skin conductance bracelet, as seen in Figure 2. Facial expression features were automatically extracted using a state-of-the-art facial expression recognition and analysis software known as FACET (commercial software previously released under the research-focused Computer Emotion Recognition Toolbox, CERT) [35]. This software provides frame-by-frame tracking of nineteen facial action units according to the Facial Action Coding System [34, 21]. These action units include expressions such as AU4 BROW LOWERER, AU15 LIP CORNER DEPRESSOR, AU18 LIP PUCKERER, and AU24 LIP PRESSOR (see Figure 5 for illustration). For each of these nineteen action units, the FACET software provides an *Evidence* measure, suggesting the chance that the target expression is present, as opposed to not present.

4. ANALYSIS AND RESULTS

The objective of the present analysis is to investigate how men and women express affect (particularly through facial expression) differently during learning. In order to identify these differences, we first examine the presence of each facial action unit in the set of female students versus the presence of each expression in the set of male students, and discuss the differences that emerge. We explore the possibility that these differences exist as a result of incoming student characteristics, but find no significant differences across genders. Finally, we examine the hypothesis that learning gain can be predicted differently by different affective features for male and female students.

4.1 Facial Expression Differences by Gender

The statistical comparisons were conducted using standard Student's *t*-tests. Additionally, a Bonferroni correction $p \leq \alpha/n$ was applied to control the familywise error rate,

where $n = 19$ is the number of statistical tests conducted (one for each facial action unit detected). The FACET Evidence measure for each action unit (see Section 3.3) was averaged across each student's session, giving $n = 67$ ratings per action unit. The results, displayed in Table 2 and illustrated in Figure 4, suggest some key differences.

In particular, female students tended to express AU15 LIP CORNER DEPRESSOR (Figure 3b) more frequently than male students ($t = -3.852$, $p = 0.008$). On the other hand, male students were more likely to express AU4 BROW LOWERER (Figure 3a, $t = +3.296$, $p = 0.015$), AU18 LIP PUCKERER (Figure 3c, $t = +4.374$, $p < 0.001$), and AU24 LIP PRESSOR (Figure 3d, $t = +3.208$, $p = 0.023$).

In order to examine whether these differences across genders may have arisen from other incoming student characteristics, we examined results from the surveys and tests administered the students prior to any interaction with the system. We detail each of these instruments below.

First, we considered the possibility that students with more incoming knowledge in computer science may be more comfortable with the material presented than those with less, and therefore may express affect differently throughout the session. We therefore examined incoming knowledge via pretest score to determine whether females and males in our study had different incoming knowledge. Prior to every lesson, students were given a content pretest on the content covered in the upcoming lesson. There was no significant difference between pretest scores between male and female students ($p > 0.99$).

Next we investigated whether the facial expression differences could have arisen from differences in self-efficacy, or how confident students are in their own ability to complete tasks and reach their goals. The higher a student's self-efficacy, the more she believes that she can succeed academically. Prior to any interaction with the tutoring environment, students completed the eight-item New General Self-

TUTOR	hang on :)
TUTOR	When we show you example code, it is not the code you need to write.
TUTOR	Look at the task again.
<i>Student writes programming code.</i>	
TUTOR	YUP
TUTOR	Perfect
TUTOR	OK. Go ahead and test.
STUDENT	And I don't need anything in the parentheses?
TUTOR	Line 9 is correct. You do NOT need anything inside the parentheses.
STUDENT	Ok
<i>Student compiles and runs code successfully.</i>	
TUTOR	Good.
TUTOR	Moving on.
<i>Tutor advances to the next task.</i>	
<i>Student writes programming code.</i>	
TUTOR	Syntactically correct. But there is a logic error
TUTOR	When will the output statement display your request to the player?
STUDENT	AFTER they put in their name
TUTOR	Exactly

Table 1: Sample dialogue between a tutor and a student occurring in the tutorial corpus.

	Females	Males
AU4 BROW LOWERER	-0.536 ± 0.418	-0.133 ± 0.566
AU15 LIP DEPRESSOR	-0.499 ± 0.296	-0.786 ± 0.281
AU18 LIP PUCKERER	-0.510 ± 0.590	0.136 ± 0.556
AU24 LIP PRESSOR	-0.353 ± 0.358	-0.051 ± 0.386

Table 2: Summary statistics (mean \pm standard deviation) for significantly different facial expressions across gender.

Efficacy scale [8]. There was no significant difference in self-efficacy scores between male and female students ($p > 0.56$)

Finally, we considered the possibility that personality traits may be influential in facial expression. We measured personality with the Big Five Factor model of personality, one of the standard frameworks for identifying personality traits [22]. Prior to any interaction with the system, students completed the 44-item Big Five Factor Inventory survey, which included items to identify Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. We hypothesized that two of these factors may have had an impact on affective expression: Extraversion and Neuroticism. Extraversion is defined as the part of the Big Five Factors that identifies gregariousness, activity, positive emotions, and warmth, whereas Neuroticism is defined as the part that identifies anxiety, impulsiveness, and negative emotions [31]. However, there was no significant difference in either Extraversion or Neuroticism scores between male and female students ($p > 0.91$).

4.2 Predicting Learning from Facial Expression

The comparative analyses described above provide evidence that men and women's facial expressions differ significantly during learning, even when incoming knowledge,

self-efficacy levels, and personality do not significantly differ across genders. Next, we examine the question of whether different facial action units are predictive of learning gains in male and female students. In order to answer this question, we built two predictive models for learning gain: one for female students and one for male students. Each feature described in Section 3.3 was standardized by subtracting the mean and dividing by the standard deviation. All of these features were then provided to a stepwise regression modeling procedure that optimizes the leave-one-student-out cross-validated R^2 value (that is, the coefficient of determination) while at the same time applying a strict $p < 0.05$ cut-off value for significance. The models are presented in Tables 3 and 4.

Table 3: Predictive model for standardized normalized learning gain in female students.

Normalized Learning Gain =	R^2	p
+0.6508 * AU2 OUTER BROW RAISER	0.2209	0.009
+1.2120 * AU9 NOSE WRINKLER	0.1109	0.003
+1.7156 * AU12 LIP CORNER PULLER	0.1591	0.006
-0.7100 * AU20 LIP STRETCHER	0.0908	0.005
-0.9414 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.5817$		

Four facial expressions are significantly predictive of learning gain in female students, three of which are positively predictive (Table 3). The more present that AU2 OUTER BROW RAISER (Figure 5a) is during the session, the more likely that the student will achieve a high learning gain at the end of the session. Similarly, the more that AU9 NOSE WRINKLER (Figure 5c) and AU12 LIP CORNER PULLER (Figure 5d) are present, the more the student tends to learn. Finally, the



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.

more present AU20 LIP STRETCHER (Figure 5e) during the session, the *less* likely that the student will achieve a high learning gain.

Table 4: Predictive model for standardized normalized learning gain in male students.

Normalized Learning Gain =	R^2	p
-0.6628 * AU5 UPPER LID RAISER	0.1199	0.010
+0.1747 (Intercept)	1.000	
Leave-One-Out Cross-Validated $R^2 = 0.1199$		

Only a single facial action unit was found to be significantly predictive of learning gain in male students (Table 4). The more present AU5 UPPER LID RAISER (Figure 5b) during the session, the more likely the student is to achieve a high learning gain at the end of the session.

4.3 Contrast Between Predictive Models of Learning

A widely known limitation of the type of stepwise regression procedure utilized here is that there can be some variation in the features selected depending on factors such as the order of addition of the predictors. In order to investigate the extent to which the two predictive models of learning differ in meaningful ways (rather than in subtle feature selection outcomes) we tested the predictive power of the female model on the data from male students, and vice versa. First, we created a comparison model for male students by providing it only the features that were significant in the female model. We then trained this comparison model on the male data. In leave-one-student-out cross-validation, the model showed no predictive power, with an R^2 of -0.29. Similarly, when a model was trained on female data but forced to use

the predictor that had performed best on male data, the R^2 was -0.04. In both comparison models, none of the facial action unit features for the other gender was significantly correlated with the learning outcome: no p -value for any feature was lower than 0.62 (p -values greater than 0.05 are not significant in the regression model).

5. DISCUSSION

A large body of evidence suggests that emotional experience, nonverbal displays of affect, and societal norms differ significantly across genders [5]. The present study has demonstrated differences in facial expression of female and male students during tutoring, as well as differing facial expression predictors of learning across genders. These results have important implications in how learning environments may adapt and personalize to students of each gender. We discuss each result and suggest specific directions for follow-up analyses.

These results reveal particular differences in lower face fidgeting of female and male students. The female students exhibited facial movements of the corners of the lips that may have been associated with moments of negative affect or uncertainty. Male students, on the other hand, exhibited more facial actions involving the lips themselves—specifically lip puckering and lip pressing. These facial movements may also have co-occurred during moments of mental effort or negative affect. Interestingly, while research on basic emotions would indicate these movements indicate negative affect [21], there is very little empirical evidence regarding these lip movements in learning contexts. Therefore, a cautious interpretation is that these results provide differential evidence of how female and male students express moments of mental effort or uncertainty during learning.

An additional facial action unit, brow lowering, was displayed more frequently by male students. This facial move-

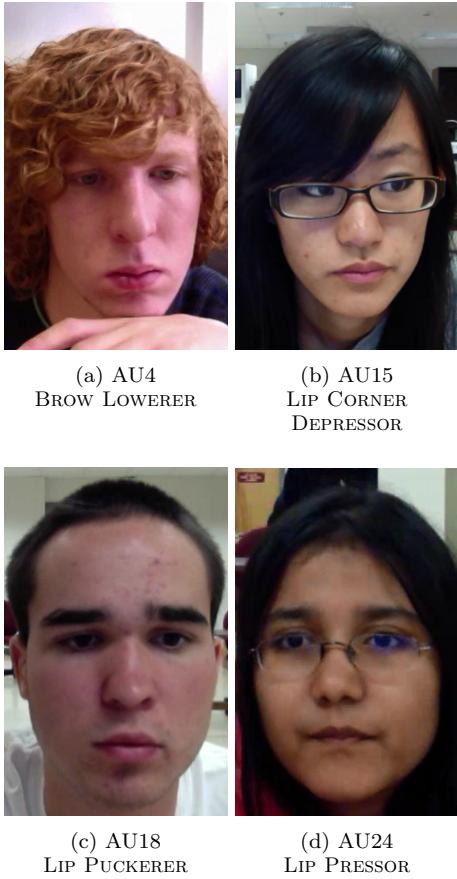


Figure 3: Sample frames from the student webcam illustrating the facial action unit features appearing significantly differently in male and female students, as identified by FACET.

ment has been acknowledged as a key indicator of mental effort or confusion in learning contexts [17, 36, 23]. This result provides evidence that male students displayed this facial movement more frequently, though it is unclear whether this corresponds to greater affective experience of confusion.

The predictive model of normalized learning gain for female students provides both positive and negative indicators of learning. Two of the positive predictors, outer brow raising (AU2) and nose wrinkling (AU9), have typically been interpreted as indicating negative facial expressions. However, as noted above, these facial movements may have distinctly different meanings in the context of learning. As a component of empirically identified anxiety, outer brow raising may occur when students are feeling overwhelmed. However, the student's subsequent progress on the learning task determines learning outcomes. Similarly, nose wrinkling moves the eyebrows in a similar way to brow lowering and has been associated with negative affect, such as anger. Again, in the context of learning, this is more likely an indicator of mental effort at salient moments of tutoring. Thus, facial movements that may naively be interpreted as negative affect are identified here as positive predictors of learning.

The remaining predictors in the model of learning for female students provide a more straightforward contrast. Both are highly social displays of emotion, with smiling

(AU12) associated with enjoyment [21] and lip stretching (AU20) identified as a component of embarrassment [32]. The evidence shows that these facial movements follow the presumed directions: smiling is predictive of higher learning gain and the component of embarrassment, lip stretching, is predictive of lower learning gain. That these socially oriented facial expressions are predictive of female students' learning seems to be consistent with the notion that females more openly and richly express emotion.

Male students, on the other hand, had a single predictor, eyelid opening (AU5), associated with lower learning gain. There are several competing interpretations for this predictor: 1) that male students opened eyes wide while being overwhelmed with the learning task, resulting in lower learning gains; 2) that widened eyes corresponded to a greater amount of reading, but this was in turn associated with extraneous cognitive processing that resulted in less efficient learning; or 3) that eye widening was performed in an attempt to stave off boredom, which has been found to be a negative affective state coinciding with lower learning gain. These competing interpretations offer the possibility of follow-up studies to determine the specific ways in which facial expressions co-occur with impactful moments during tutoring.

Notably, the predictive models of learning included facial movements that were entirely separate from the comparisons of facial expression frequency across the genders. Female students were found to more frequently display lip corner fidgeting, while male students puckered and pressed their lips more frequently. Additionally, male students lowered their brows more frequently, which has been noted as a prominent indicator of mental effort in past research. As these facial action units were not predictive of learning for each gender, they appear to indicate more subtle nuances of facial expression across female and male students. It is unclear whether societal gender roles would produce these differences, but automated facial expression recognition provides a tool to explore how these occur within the context of learning. These results have highlighted gender differences in facial expression that are both predictive of learning and show how nonverbal behavior associated with internal states relevant to learning (e.g., mental effort) may be expressed differently across genders.

6. CONCLUSION

Modeling users' affect during learning is one of the central challenges for user-adapted systems. Of all the channels through which affect can be detected, facial expression offers one of the richest. However, there is significant evidence that the expression of emotions through facial expression differs substantially across genders, with women typically expressing in a somewhat more pronounced way. However, the ways in which facial expressions during learning differ between genders is an open question.

We have presented an analysis of a data set of one-on-one computer-mediated human tutoring in which we collected students' facial expressions, automatically tagged them with facial expression analysis software, and compared the occurrence of all facial action units across genders. We found significant differences, with females displaying one lower face movement around the mouth more frequently than men, while men displayed more brow lowering and different lower face movements. These findings suggest that women and

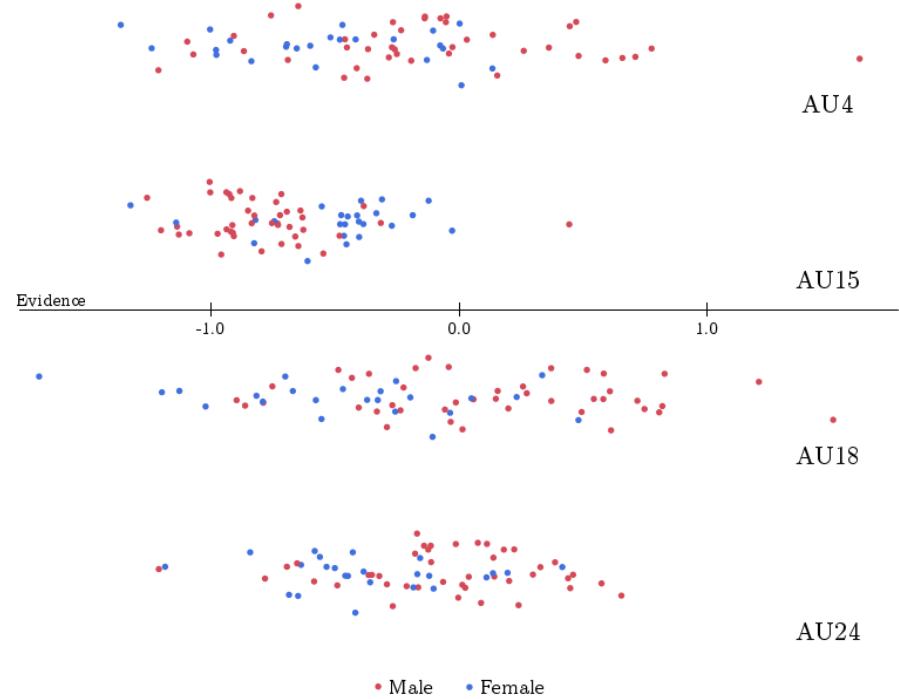


Figure 4: Distribution of average Evidence measure of each of the facial expressions found to be expressed differently between male and female students

men express moments of mental effort or uncertainty differently during tutoring. We also found that indicators of effort and components of embarrassment were predictive of women’s learning outcomes, while eyelid opening negatively predicted men’s learning outcomes. We suggest three potential explanations as to why this finding would have emerged, which in future work should be explored to elucidate the causes.

These results point to several important directions for future work. First, it is important to gain a deeper understanding of the affective processes that underlie the observations reported here. Triangulating different modalities in order to understand how the facial expressions displayed by learners co-occur with physiological indicators of emotion will shed light on the affective processes themselves. Second, it is important to investigate adaptive strategies that respond to detected affective states, via either facial expression or other channels. It is hoped that this line of investigation will lead to a more complete understanding of learning-centered affective processes, reveal how they differ across genders, and lead to gender-sensitive user-adaptive learning support.

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 DRL-1007962, IIS-1409639, CNS-1042468, 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] I. Arroyo, W. Burleson, T. Minghui, M. K, and B. P. Woolf. Gender Differences in the Use and Benefit of Advanced Learning Technologies for Mathematics. *Journal of Educational Psychology*, 105(4):957–969, 2013.
- [2] I. Arroyo, D. G. Cooper, W. Burleson, B. P. Woolf, M. K, and R. M. Christopherson. Emotion Sensors Go To School. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education*, pages 17–24, 2009.
- [3] I. Arroyo, B. P. Woolf, D. G. Cooper, W. Burleson, and K. Muldner. The impact of animated pedagogical agents on girls’ and boys’ emotions, attitudes, behaviors and learning. In *Proceedings of the 11th International Conference on Advanced Learning Technologies*, pages 506–510, 2011.

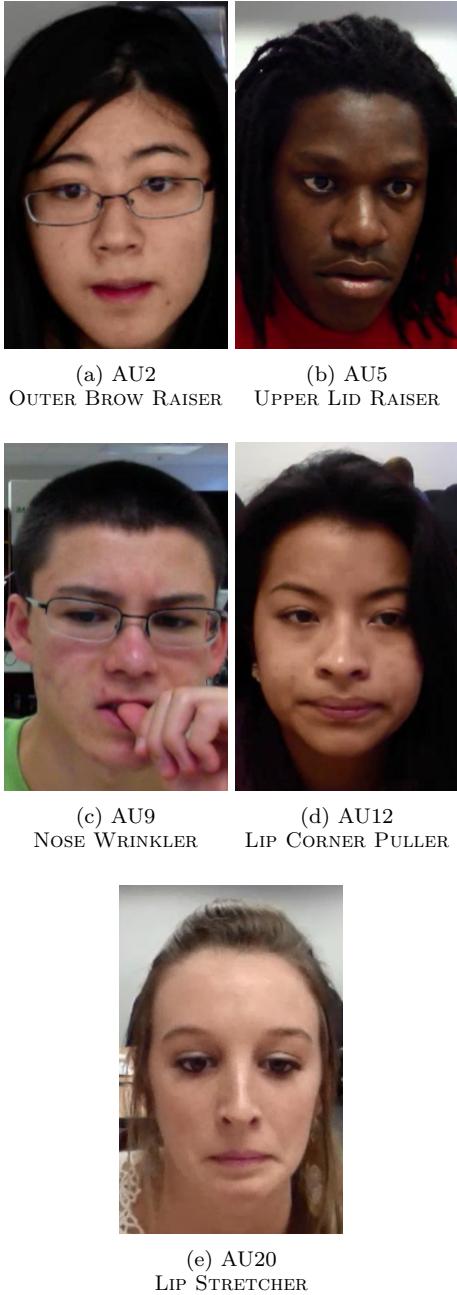


Figure 5: Sample frames from the student webcam illustrating the facial action unit features appearing significantly in predictive models of learning gain in male and female students, as identified by FACET.

- [4] R. S. J. d. Baker, S. K. D'Mello, M. T. Rodrigo, and A. C. Graesser. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4):223–241, 2010.
- [5] L. R. Brody and J. A. Hall. Gender and emotion in context. *Handbook of emotions*, 3:395–408, 2008.
- [6] W. Burleson and R. W. Picard. Gender-specific approaches to developing emotionally intelligent learning companions. *IEEE Intelligent Systems*, 22(4):62–69, 2007.
- [7] R. A. Calvo and S. K. D'Mello. *New Perspectives on Affect and Learning Technologies*. Springer Science & Business Media, 2011.
- [8] G. Chen, S. M. Gully, and D. Eden. Validation of a new general self-efficacy scale. *Organizational Research Methods*, 4(1):62–83, 2001.
- [9] M. Chi, K. VanLehn, D. J. Litman, and P. W. Jordan. Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. *User Modelling and User-Adapted Interaction*, 21(1-2):137–180, 2011.
- [10] S. Corrigan, T. Barkley, and Z. Pardos. Dynamic Approaches to Modeling Student Affect and its Changing Role in Learning and Performance. In *Proceedings of the 23rd International Conference on User Modeling, Adaptation, and Personalization*, pages 92–103, 2015.
- [11] S. Craig, S. D'Mello, A. Witherspoon, and A. Graesser. Emote-Aloud during Learning with AutoTutor: Applying the Facial Action Coding System to Cognitive-Affective States during Learning. *Cognition and Emotion*, 22(5):777–788, 2008.
- [12] S. D. Craig, A. C. Graesser, J. Sullins, and B. Gholson. Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3):241–250, 2004.
- [13] F. De la Torre and J. F. Cohn. Facial expression analysis. *Visual analysis of humans*, pages 377–409, 2011.
- [14] M. Dennis, J. Masthoff, and C. Mellish. Adapting Performance Feedback to a Learner's Conscientiousness. In *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization*, pages 297–302, 2012.
- [15] M. C. Desmarais and R. S. J. d. Baker. A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modelling and User-Adapted Interaction*, 22(1-2):9–38, 2012.
- [16] U. Dimberg and L.-O. Lundquist. Gender differences in facial reactions to facial expressions. *Biological psychology*, 30(2):151–159, 1990.
- [17] S. D'Mello, B. Lehman, R. Pekrun, and A. Graesser. Confusion can be beneficial for learning. *Learning and Instruction*, 29:153–170, 2014.
- [18] S. K. D'Mello. A Selective Meta-analysis on the Relative Incidence of Discrete Affective States during Learning with Technology. *Journal of Educational Psychology*, 105(4):1082–1099, 2013.

- [19] S. K. D'Mello and A. C. Graesser. AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, 2:1–39, 2012.
- [20] P. Ekman, W. V. Friesen, and P. Ellsworth. *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Pergamon Press, 1972.
- [21] P. Ekman, W. V. Friesen, and J. C. Hager. *Facial Action Coding System: Investigator's Guide*. 2002.
- [22] L. R. Goldberg. The structure of phenotypic personality traits. *American Psychologist*, 48(1):26–34, 1993.
- [23] 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*, pages 97–106, 2011.
- [24] J. F. Grafsgaard, S. Y. Lee, B. W. Mott, K. E. Boyer, and J. C. Lester. Modeling Self-Efficacy Across Age Groups with Automatically Tracked Facial Expression. In *Proceedings of the 17th International Conference on Artificial Intelligence in Education*, pages 582–585, 2015.
- [25] 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*, pages 43–50, 2013.
- [26] J. F. Grafsgaard, J. B. Wiggins, A. K. Vail, K. E. Boyer, and J. C. Lester. The Additive Value of Multimodal Features for Predicting Engagement, Frustration, and Learning during Tutoring. In *Proceedings of the 16th ACM International Conference on Multimodal Interaction*, pages 42–49, 2014.
- [27] D. T. Green, T. J. Walsh, P. R. Cohen, C. R. Beal, and Y.-H. Chang. Gender Differences and the Value of Choice in Intelligent Tutoring Systems. In *Proceedings of the 19th International Conference on User Modeling, Adaptation, and Personalization*, pages 341–346, 2011.
- [28] 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.
- [29] J. A. Hall. Gender effects in decoding nonverbal cues. *Psychological bulletin*, 85(4):845, 1978.
- [30] J. A. Hall and D. Matsumoto. Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, 4(2):201, 2004.
- [31] O. P. John and S. Srivastava. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, 2:102–138, 1999.
- [32] D. Keltner. Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, page 441.
- [33] M. LaFrance and M. A. Hecht. Option or obligation to smile: The effects of power and gender on facial expression. *The social context of nonverbal behavior: Studies in emotion and social interaction*, page 431, 1999.
- [34] 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.
- [35] 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 11th International Conference on Automatic Face & Gesture Recognition and Workshops*, pages 298–305, 2011.
- [36] G. C. 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 International Conference on Automatic Face & Gesture Recognition and Workshops*, pages 30–35, 2011.
- [37] 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.
- [38] A. Mitrovic. Fifteen years of constraint-based tutors: What we have achieved and where we are going. *User Modelling and User-Adapted Interaction*, 22(1-2):39–72, 2012.
- [39] M. O. Z. San Pedro, R. S. J. d. Baker, S. M. Gowda, and N. T. Heffernan. Towards an Understanding of Affect and Knowledge from Student Interaction with an Intelligent Tutoring System. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, pages 41–50, 2013.
- [40] A. K. Vail, K. E. Boyer, E. N. Wiebe, and J. C. Lester. The Mars and Venus Effect: The Influence of User Gender on the Effectiveness of Adaptive Task Support. In *Proceedings of the 23rd International Conference on User Modeling, Adaptation, and Personalization*, pages 216–227, 2015.
- [41] K. VanLehn. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221, 2011.
- [42] K. VanLehn, A. C. Graesser, G. T. Jackson, P. W. Jordan, A. Olney, and C. P. Rosé. When Are Tutorial Dialogues More Effective Than Reading? *Cognitive Science*, 31(1):3–62, 2007.
- [43] M. Wixon and I. Arroyo. When the Question is Part of the Answer: Examining the Impact of Emotion Self-reports on Student Emotion. In *Proceedings of the 22nd International Conference on User Modeling, Adaptation, and Personalization*, pages 471–477, 2014.
- [44] B. P. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. G. Cooper, and R. W. Picard. Affect-Aware Tutors: Recognizing and Responding to Student Affect. *International Journal of Learning Technology*, 4(3–4):129–164.