Predicting Learning and Engagement in Tutorial Dialogue: A Personality-Based Model

Alexandria Katarina Vail, Joseph F. Grafsgaard, Joseph B. Wiggins,
James C. Lester, Kristy Elizabeth Boyer
Department of Computer Science
North Carolina State University, Raleigh, North Carolina, USA
{akvail, jfgrafsg, jbwiggi3, lester, keboyer}@ncsu.edu

ABSTRACT
A variety of studies have established that users with different personality profiles exhibit different patterns of behavior when interacting with a system. Although patterns of behavior have been successfully used to predict cognitive and affective outcomes of an interaction, little work has been done to identify the variations in these patterns based on user personality profile. In this paper, we model sequences of facial expressions, postural shifts, hand-to-face gestures, system interaction events, and textual dialogue messages of a user interacting with a human tutor in a computer-mediated tutorial session. We use these models to predict the user's learning gain, frustration, and engagement at the end of the session. In particular, we examine the behavior of users based on their Extraversion trait score of a Big Five Factor personality survey. The analysis reveals a variety of personality-specific sequences of behavior that are significantly indicative of cognitive and affective outcomes. These results could impact user experience design of future interactive systems.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human factors, Human information processing; H.5.2 [**User Interfaces**]: Natural language; J.4 [**Social and Behavioral Sciences**]: Psychology

General Terms
Experimentation, Human Factors

Keywords
Personality; Tutorial Dialogue; Engagement; Frustration; Learning

1. INTRODUCTION
Multimodal learning analytics is a growing research area that focuses on multiple communication modalities, including speech, writing, posture, gesture, and facial expressions [28]. In the field of learning analytics, the goal is to predict student learning and learning-related outcomes in a reliable and effective manner. Multimodal learning analytics expands upon this mission by supporting a richly featured view of the learning process, allowing for a deeper understanding of how students communicate, collaborate, and grow to understand material while learning. In particular, multimodal learning analytics have the potential to help create systems that are more reliable and more adaptive than prior work. This paper explores multimodal analytics for supporting one of the most effective modes of human learning: tutorial dialogue [7, 10].

One promising direction for creating more personalized tutoring environments is to investigate the affective channel, a key component of natural 'human' interaction [13, 26]. Understanding this affective channel and its implications for goal-oriented interactions is a promising direction for improvement in effective systems that aim to harness some of the power of human communication. In particular, the study of nonverbal displays of affect, such as facial expression, posture, and gesture is receiving increasing attention [8]. Examining these features holds great promise for capturing a rich picture of the learning interactions.

Another important direction for developing adaptive systems is to incorporate and adapt based upon incoming user traits. While incoming traits such as knowledge level have long been considered central, it has been suggested by some studies that users of different personalities may react more favorably to different system behaviors [27, 30]. Prior work has also demonstrated that this benefit may extend to the context of learning [31]. Some studies have suggested that extraverted or introverted tendencies of the user are particularly sensitive to system behavior, a critical concern when designing tutorial systems [5].

We hypothesize that the cognitive and affective outcomes of user interactions can be better predicted when incorporating incoming student traits. We utilize a widely validated personality questionnaire, the Big Five Factor Inventory, to determine the personality profile of each student [11]. We then split the students into 'introverts' and 'extraverts', and build predictive models of three cognitive and affective outcomes of the interaction: learning gain, frustration, and engagement. The results suggest that predictive models built on classes of student personality are more effective than those that are not. In particular, we discover that extraverts benefit significantly from some tutorial dialogue moves, whereas introverts may not be. These results could benefit the design of the next generation of adaptive learning environments.

2. RELATED WORK

Most studies to date have examined the use of verbal and nonverbal behaviors to predict user personality, particularly in conversational video. One line of investigation has examined YouTube video blogs with the goal of automatically predicting blogger *personality impression* on the Big Five Factor personality scale. Unlike the present analysis, personality impressions involve an external observer completing the survey, rather than the user completing the survey himself. Several conversational video blogs were downloaded from YouTube and characterized with the Computer Expression Recognition Toolbox (CERT) universal facial expressions of emotion: 'anger', 'joy', 'contempt', 'disgust', 'surprise', 'sad', 'fear', and 'neutral'. Significant success has been found in this domain when attempting to predict personality impressions of Extraversion, Conscientiousness, and Openness, in particular [5]. In an expanded follow-up survey, the Extraversion trait was found to be most accurately predicted by facial indicators of emotion among all five traits [4].

A further study on the same corpus of YouTube video blogs examined verbal content cues for indicators of personality. The analysis revealed that the content of student utterances can significantly improve the accuracy of prediction algorithms for user scores on the Big Five Factor personality traits in general [6]. However, the improvement in Extraversion prediction was not statistically significant; the authors suggest that a user's Extraversion trait would be best predicted using audiovisual activity and facial emotion cues, rather than dialogue content.

A similar analysis examined group meetings, in which three or four persons were asked to debate preparation for a hypothetical situation. Audio and video were captured from all members of the group. It was revealed that Extraversion was the best predicted trait in this domain as well, producing a prediction algorithm with a 75% accuracy. The selected features suggest that users scoring high in Extraversion exhibited more energy in their nonverbal movements than those that score lower [2]. A subsequent cross-domain analysis attempted to apply the previous YouTube video blog model to the group meeting corpus, and discovered that Extraversion impressions could still be predicted with 70% accuracy [1]. This result suggests that the audiovisual cues indicative of of Extraversion could be generalizable across domains.

Although these studies have suggested promising directions for the prediction of Extraversion in users, it is important to question whether similar reflections of personality can be discovered in the educational domain. Some studies have examined groups of students working through a set of problems in mathematics, attempting to predict incoming student traits with a rich multimodal dataset containing digital paper notes, audio recordings, multi-angle video recordings, and writing patterns of the students. These studies attempted to identify two student traits in particular: social dominance (the 'leader' quality) and domain knowledge (the 'expert' quality). A preliminary study revealed a variety of nonverbal cues specific to these types of students, including the energy level of the user's voice and the ratio between time spent writing and time spent speaking [28]. This study was followed by an analysis of the same corpus with the goal of predicting solution correctness in advance of the actual submission. It was discovered that the 'leader' and 'expert' rank-order of the students had a significant impact on the solution correctness prediction algorithm [25]. This finding suggests that user personality traits are pertinent to the prediction of student learning.

In contrast to these prior studies, which attempted to use expressed affect and learning to predict personality traits, the present analysis takes into account a student's Extraversion score before attempting to predict learning and affective outcomes. Prior work has discovered key differences in dialogue progression between tutoring sessions with extraverts and those with introverts [31]. We hypothesize that this difference extends to nonverbal behaviors as indicators of mental state. That is, we hypothesize that students of differing Extraversion tendencies will exhibit different behaviors to indicate the same cognitive and affective states, and we attempt to model these differences through models that predict learning.

3. DATA COLLECTION

The corpus examined in this analysis consists of computer-mediated human-human interactions during tutorial sessions in introductory computer science (specifically, programming in Java) [23, 17]. Each tutoring session consisted of an interaction between one tutor and one student collaborating to create a text-based adventure game during a series of six lessons. Participants were university students in the United States (average age 18.5 years, $s = 1.5$ years) who participated voluntarily in exchange for course credit in an introductory engineering course. No previous computer science knowledge was assumed or required. Tutors were primarily graduate students with prior experience in tutoring or teaching Java programming.

Students and tutors interacted through a web-based interface for introductory programming in Java, shown in Figure 1.The interface consists of four panes: the task objective description, the student's Java source code, the compilation and execution output, and the textual dialogue messages between the tutor and the student. The student could write, compile, and execute source code, as well as send textual dialogue messages to the remote tutor. The interface was synchronized in real time between the tutor and the student; however, the tutor's interactions with the system were limited to sending textual messages to the student and displaying the next task when the student completed the current objective.

The data were collected during the Fall 2011 and Spring 2012 academic semesters. During this time frame, $N = 67$ student-tutor pairs completed the entire set of six lessons over four weeks, each lesson constrained to forty minutes. Data recorded from the interactions consisted of database logs, webcam video, Kinect depth video, and skin conductance readings. This study examines the database logs, webcam video, and Kinect depth video from the first lesson only. This first lesson covered the development of a text-based adventure, introducing simple concepts such as conditionals and variables.

Prior to each tutorial session, students were administered a content-based pretest. After the completion of the session, students completed a posttest (identical to the pretest) and a post-session survey. The post-session survey included the User Engagement Survey [24] and the NASA-TLX workload

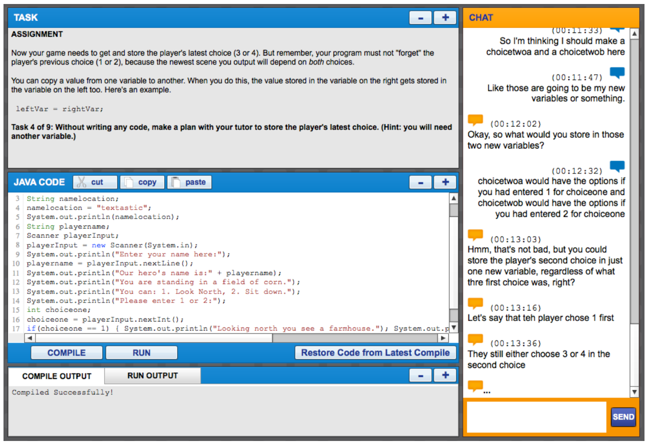


Figure 1: The web-based interface for introductory Java programming.

survey [18], which included an item for Frustration Level. We focused on Frustration level in particular, as prior work has shown that Frustration has a significant impact on learning [3]. These are the cognitive and affective outcomes we aim to predict in the present analysis (see Section 5).

3.1 Personality Profile Measurement

Prior to the first tutoring session, students completed a pre-survey constructed to measure incoming student characteristics. Included in this survey was a questionnaire developed for identification of five personality traits [12, 19]. This model of personality includes items to identify extraversion vs. intraversion [11]. The overall questionnaire included 44 items created to place an individual on a scale for five different factors of personality: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [11], but the analysis detailed in this paper focused only on the items pertaining to the Extraversion trait, identified in Table 1. According to this model, the Extraversion score measures a person's assertiveness, activity, gregariousness, positive emotions, excitement-seeking nature, and warmth [19].

I see myself as someone who ...
... is talkative.
... is reserved.*
... is full of energy.
... generates a lot of enthusiasm.
... tends to be quiet.*
... has an assertive personality.
... is sometimes shy, inhibited.*
... is outgoing, sociable.

Table 1: Items of the Big Five Inventory used to identify a person's Extraversion score. Asterisks (*) represent items negatively associated with extraversion.

The possible Extraversion score on the questionnaire ranges from -10 (highly introverted) to 25 (highly extraverted). The mean Extraversion score of the students in this corpus was 6.40 ($s = 6.42$). The distribution of scores was largely normally distributed. Students were binned into two groups: the 'introverts', who scored below or equal to the median extraversion score of 7, and the 'extraverts', who scored above the median score. These groups contained 34 students and 33 students, respectively. We split on the median score of our student sample rather than a larger population median because personality ranges greatly by sample, and no large study to date has examined university students to establish personality norms.

4. MULTIMODAL CORPUS

Data from the Kinect depth camera, computer webcam, and database logs were combined to form the multimodal corpus examined in this analysis. The features generated from each data stream are detailed in the following subsections.

4.1 Gestural and Postural Features

The Kinect depth camera video was used to discover posture and gesture of the student during the tutorial session. Previously-developed tracking algorithms were employed to identify these features; prior evaluations of these algorithms determined an accuracy of 92.4% in posture tracking and 92.6% in gesture tracking [13]. Gestures consisted of one- or two-hand-to-face gestures (OneHTF and TwoHTF, respectively). With regards to posture, the average student distance from the workstation was denoted the 'mid' postural position; distances at one or more standard deviations closer or farther from 'mid' were considered 'near' or 'far', respectively. The events MoveForward and MoveBack were extracted from student movements between these postural zones.

4.2 Facial Expression Features

Identification of facial expressions was performed by a state-of-the-art facial expression recognition tool, the Computer Expression Recognition Toolbox (CERT) [21]. Providing frame-by-frame tracking of a wide variety of facial action units by locating faces in each frame, CERT identifies facial features for the nearest face and calculates weights for each facial action unit using support vector machines [20]. This analysis was applied to the webcam video collected during the first tutoring session.

Five action units were selected for inclusion in the present study, based on a validation of adjusted CERT output against manual FACS annotations in prior work on tutorial dialogue [14]. The adjustment to the CERT output consisted subtracting the average value from each measurement, as a baseline, to adjust for systematic tracking error. Furthermore, although CERT considers any positive output value an action unit, it was empirically determined that a higher threshold results in a reduction of false positives; thus, an an action unit was recorded when the baseline-adjusted CERT output was equal to or greater than $\lambda = 0.25$. One of five possible events was recorded when a facial action unit was triggered: AU1 (Inner Brow Raise), AU2 (Outer Brow Raise), AU4 (Brow Lowering), AU7 (Lid Tightener), and AU14 (Mouth Dimpling).

4.3 Task Event Features

As students progressed through each task, the system logged dialogue messages, typing status, and task progress to the database. Tutorial dialogue occurred at any time during the session; no turn-taking was enforced (see Subsection 4.4 for details). When the student began writing code, the system logged a typing event (CODING). After the student finished coding, he could attempt to compile the code (COMPILESTART), which would either be successful (COMPILESUCCESS) or encounter an error (COMPILEERROR). A student could also run the code (RUNPROGRAM). Finally, when the student stopped interacting with the system for over three seconds (a manually-determined window), a pause was registered (TASKPAUSE).

4.4 Dialogue Features

All student and tutor messages in the corpus were tagged with a previously-developed dialogue act annotation scheme [32]. This annotation scheme represents a refinement of previous dialogue act tagsets developed for task-oriented tutoring [17], with an emphasis on the decomposition of frequent tags, in order to capture more fine-grained interactions in the dialogue. The tags included in this annotation scheme are listed in Table 2.

We utilize an automated decision tree-based classifier trained on a manually tagged subset of the corpus [31]. This classifier achieved an accuracy of 80.11% (Cohen's kappa of 0.786) on a held-out test set tagged manually. The present analysis makes use of these automatically-tagged dialogue utterances. Each tutor message was recorded as one event, the sending of the dialogue message (e.g., POSFEEDBACK (Tutor)), but each student message was recorded with timestamp for the start of typing the message, e.g., TYPINGACK, and the sending of the message, e.g., ACKNOWLEDGE (Student). This additional logging is so that multimodal features co-occurring with the formulation of a message can be identified.

5. ANALYSIS

The goal of the present analysis is to identify differences in student behavior between extraverts and introverts and to determine whether these differences are predictive of tutoring outcomes. In particular, we examine the conditional probabilities of any set of two events occurring in sequence (i.e., the probabilities $Pr(X_t|X_{t-1})$) within a three-second interval during each tutorial session. We had initially performed an exploratory analysis of isolated unigrams, but none were significant. Each row of the data set to be modeled represents the conditional probability features summarizing a single student's tutorial session.

Initially, all variables were standardized (centered at the mean and scaled to a unit standard deviation) to enable comparison between variables. Feature selection was then performed using the model averaging feature of the JMP statistical software (constructing all models of one or two features, and computing the average coefficient magnitude of each variable across all models) [29]. The top 10% of features were selected using the average coefficient magnitude estimate from models with one or two predictive variables. The predictive models described in the following sections were constructed with the goal of maximizing the leave-one-out cross-validated R^2 value (the coefficient of determination), while enforcing a strict $p < 0.05$ cut-off for significance of included variables. Separate models were built for extraverted and introverted students for each dependent variable: learning gain, frustration, and engagement. These models are described in the following subsections.

5.1 Predicting Learning Gain

Normalized learning gain was computed using pretest and posttest scores, as in Equation 1. This equation measures how much a student learned relative to how much she could have learned, based on her pretest score [22].

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

Using the feature selection approach described above, the predictive model for learning gain in extraverts contains three features. The more likely an inner brow raise (AU1) was followed by positive feedback from the tutor (POSFEEDBACK (Tutor)), or the more frequently a student moved forward (MOVEFORWARD) after an outer brow raise (AU2), the higher the student's learning gain at the end of the session. The former result may emerge from parallel reactions to a successful event: the student expresses a nonverbal indicator of joy, and the tutor congratulates the student on his work. A negative correlation with learning gain was observed when positive tutor feedback (POSFEEDBACK (Tutor)) was more frequently followed by an outer brow raise (AU2). This sequence may result from a student disbelieving the positive feedback, perhaps because he does not understand the material as well as the tutor perceived. The cross-validated model effect size (the measure of variance explained by the model) was $r = 0.6183$. The model is shown in Table 3, and selected features are illustrated in Figure 2a.

Students Tending Toward Extraversion

Normalized Learning Gain =	R^2	p
-0.5012 * POSFEEDBACK (Tutor) → AU2	0.1952	0.003
0.3574 * AU1 → POSFEEDBACK (Tutor)	0.1053	0.007
0.2877 * AU2 → MOVEFORWARD	0.0818	0.024
-0.0420 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.3823$		

Table 3: Predictive model for standardized normalized learning gain in students scoring above the median in Extraversion.

The predictive model for learning gain in introverts contains five features, the largest model built in this analysis. Two features were purely postural: a movement backward through two postural 'zones' (MOVEBACK → MOVEBACK) and a forward movement followed by a movement backward (MOVEFOWARD → MOVEBACK). Both of these features were negatively correlated with learning gain. Qualitative examination of the video corpus revealed that these sequences were frequently activated by a student shifting in his seat, which suggests discomfort, particularly in introverts, as extraverts are naturally more restless [2]. The more frequently that the composition of an acknowledgement (TYPINGACK) was followed by eyelid tightening (AU7), or the more likely that eyelid tightening (AU7) was followed by a two-hands-to-face gesture (TWOHTF), the lower the predicted learning gain. Since student acknowledgements are usually given following a tutor explanation or information turn, the former may indicate that the student is skeptical of the information that the tutor provided; introverts are more likely to simply acknowledge than question the information and engage in deeper dialogue [19]. Finally, a mouth dimpling (AU14) frequently followed by an outer brow raise

Dialogue Acts			
Acknowledge (ACK)	*Okay.*	Greeting (GRE)	*Have a good day!*
Extra Domain Answer (AEX)	*I'm doing well.*	Information (I)	*Variable names must be one word.*
Ready Answer (AR)	*Yes, I'm ready.*	Observation (O)	*See, we have an error.*
WH-Question Answer (AWH)	*Line 9.*	Extra Domain Other (OEX)	*Calculus is difficult.*
Yes/No Answer (AYN)	*No, sir.*	Confirmation Question (QC)	*It's line 6, right?*
Correction (CO)	**explanation*	Direction Question (QD)	*What do I do next?*
Directive (D)	*Test your program.*	Evaluative Question (QE)	*Does that make sense?*
Explanation (E)	*Your code stops on line 2.*	Extra Domain Question (QEX)	*How are you today?*
Negative Feedback (FN)	*No, that's incorrect.*	Factual Question (QF)	*What line is it waiting on?*
Elaborated Negative Feedback (FNE)	*That's not the right syntax.*	Information Question (QI)	*Why does that happen?*
Not Understanding Feedback (FNU)	*I don't know why that works...*	Open Question (QO)	*How can you fix it?*
Other Feedback (FO)	*That's an okay implementation.*	Probing Question (QP)	*Do you think that looks correct?*
Elaborated Other Feedback (FOE)	*That's alright, but you need to fix line 9.*	Question Prompt (QQ)	*Any questions?*
Positive Feedback (FP)	*Very good!*	Ready Question (QR)	*Ready to move on?*
Elaborated Positive Feedback (FPE)	*That's a very good approach.*	Reassurance (R)	*We have plenty of time left.*
Understanding Feedback (FU)	*Oh, that makes sense!*		

Table 2: Dialogue act tags.

was negatively associated with learning gain. The cross-validated model effect size was $r = 0.7593$. The model is shown in Table 4, and selected features are illustrated in Figure 2b.

Students Tending Toward Introversion

Normalized Learning Gain =	R^2	p
-0.8210 * MoveBack → MoveBack	0.0896	0.001
-0.4717 * AU14 → AU2	0.3051	0.009
-0.4212 * AU7 → TwoHTF	0.0840	0.017
-0.3912 * MoveForward → MoveBack	0.0430	0.033
-0.4159 * TypingAck → AU7	0.0548	0.047
-0.2116 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.5765$		

Table 4: Predictive model for standardized normalized learning gain in students scoring below the median in Extraversion.

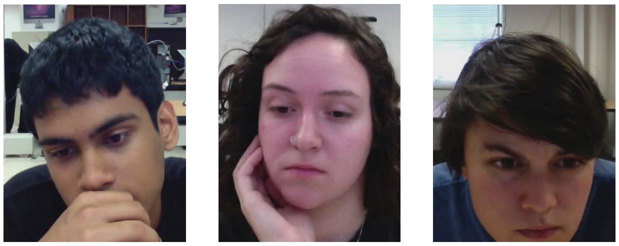
5.2 Predicting Frustration Level

The Frustration Level scale of the NASA-TLX workload survey consists of a student's self-report of how insecure, discouraged, irritated, or annoyed she felt during the session [18]. Interestingly, no statistically significant predictive model could be built for Frustration Level in students scoring below the median in Extraversion.

The predictive model of frustration for extraverts was composed of four features. A higher likelihood that a movement backward (MoveBack) was followed by an inner brow raise (AU1) or mouth dimpling (AU14) was correlated with lower frustration in the student. Mouth dimpling (AU14) has been correlated with learning gain in a prior study on the same corpus [15]. A mouth dimpling (AU14) following a backward movement (MoveBack) suggests a confident student, relaxing in her seat while working through the task. Extraverts tend to express more energy in postural movement [2], so an exaggerated motion such as this may be more likely to be observed with an extravert. Similarly, if a student was more likely to mouth dimple (AU14) after eyelid tightening (AU7), the less likely that she was frustrated. Interestingly, the final feature, a hand-to-face gesture (OneHTF) following a successful compilation (CompileSuccess), was negatively correlated with frustration in this analysis, and has been correlated with higher engagement in a prior study on the same corpus [16]. This suggests that when extraverts are more engaged with the system, they are less frustrated; this agrees with the fact that extraverts are more likely to enjoy engagement [19]. The cross-validated model effect size was $r = 0.8044$. The model is shown in Table 5, and selected features are illustrated in Figure 3.

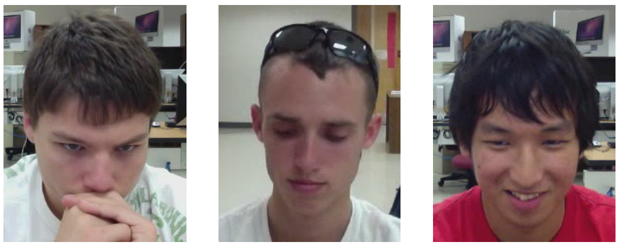
5.3 Predicting Engagement

Each student's engagement score was computed as the sum of the Focused Attention (perception of time passing), the Felt Involvement (perception of being involved in the task), and the Endurability (perception of the experience as worthwhile) sub-scales of the User Engagement Survey administered at the end of each session [24].



AU1 → PosFdbk (Tutor) | AU2 → MoveFwd | PosFdbk (Tutor) → AU2

(a) Features included in the predictive model for normalized learning gain in extraverts.



AU7 → TwoHTF | AU14 → AU2 | TypingAck → AU7

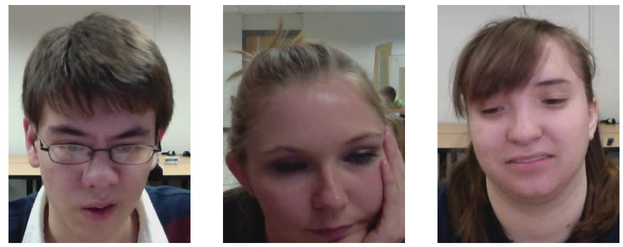
(b) Features included in the predictive model for normalized learning gain in introverts.

Figure 2: Selected features of the predictive models built on student learning gain.

Students Tending Toward Extraversion

Frustration Level =	R^2	p
−0.6437 * MoveBack → AU1	0.1528	< 0.001
−0.6346 * AU7 → AU14	0.2417	< 0.001
−0.3698 * CompileSuccess → OneHTF	0.1597	0.003
−0.4220 * MoveBack → AU14	0.0928	0.004
0.1203 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.6470$		

Table 5: Predictive model for standardized Frustration Level in students scoring above the median in Extraversion.



AU7 → AU14 | C. Success → OneHTF | MoveBack → AU14

Figure 3: Features included in the predictive model for Frustration Level in extraverts.

The predictive model for engagement in extraverts contains four features. This model is more reliant on dialogue features than any other model in the analysis. The more likely a brow lowering (AU4) was followed by an inner brow raise (AU1) or a pause in task activity (TaskPause), the lower the perceived engagement by the student. An inner brow raise (AU1) more frequently followed by an explanation from the tutor (Explanation (Tutor)) was similarly negatively correlated with engagement. It is possible that this may have been an unnecessary tutor intervention; the student encountered difficulty in the task, but the tutor stepped in to explain the problem before the student could work through the problem herself. This would agree with the interpretation of an inner brow raise (AU1) as an indicator of surprise or anxiety. The only feature positively associated with engagement in extraverts was the likelihood of an inner brow raise (AU1) after an information turn from the tutor (Information (Tutor)). Considering that an information turn is usually unprompted extra information, this may indicate that the student was surprised by the new tutor move; as before mentioned, inner brow raise (AU1) has been previously associated with surprise. The cross-validated model effect size was $r = 0.6630$. The model is shown in Table 6, and selected features are illustrated in Figure 4a.

The predictive model for engagement in introverts contains four features. Brow lowering (AU4) was prominently featured in this model; if it is likely to follow an eyelid tightening (AU7), it suggests less engagement, but if it is likely to follow the typing of an acknowledgement (TypingAck), it suggests more engagement. Recall that eyelid tightening (AU7) also featured prominently in the learning gain model for introverts, negatively in both cases. This may suggest that a student does not understand the material, and is therefore becoming less engaged with the system. Both of these interpretations align with the prototypical definition of introverts, who are less likely to seek extra discussion for additional information [19]. In the case of the other feature, as before mentioned, student acknowledgements frequently followed tutor explanation or information turns, so brow lowering (AU4) may be a nonverbal sign that the student is considering the information that the tutor provided and is in the process of internalizing the new material; brow lowering (AU4) has been previously associated with student struggle. The other two features involved task actions. The more likely an unsuccessful compilation (CompileError) follows a mouth dimpling (AU14), the lower the predicted engagement of the student. The more likely a student pauses (TaskPause) after moving forward (MoveForward), the higher the predicted engagement. The cross-validated model effect size was $r = 0.7572$. The model is shown in Table 7, and selected features are illustrated in Figure 4b.

6. DISCUSSION

Understanding how multimodal displays are associated with personality and the outcomes of tutorial dialogue holds great promise for understanding and enhancing the user experience with adaptive learning environments. We have presented predictive models of learning, engagement, and frustration that selected features from among a variety of dialogue, task, and multimodal data streams. The results show that predictive models for extraverts tend to rely on dialogue, particularly from the tutor, more often than the predictive models for introverts. Although extraverts by definition engage in increased conversation compared to introverts [19], it is interesting to note that dialogue has a meaningful impact on tutorial outcomes as well. In particular, student engagement and learning gain were discovered to be positively or negatively affected by certain dialogue moves on the part of the tutor. Frustration, however, was more often

Students Tending Toward Extraversion

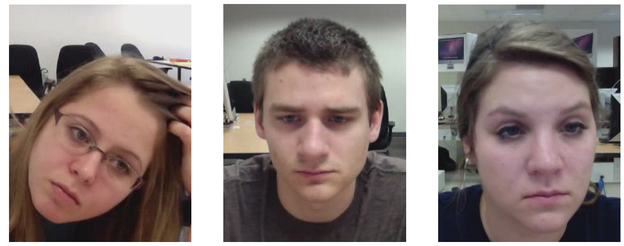
Engagement =	R^2	p
-0.5166 * AU4 $\rightarrow$ TaskPause	0.2196	< 0.001
0.2976 * Information (Tutor) $\rightarrow$ AU1	0.0925	0.017
-0.3976 * AU1 $\rightarrow$ Explanation (Tutor)	0.0420	0.008
-0.2904 * AU4 $\rightarrow$ AU1	0.0855	0.019
-0.2592 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.4396$		

Table 6: Predictive model for standardized engagement in students scoring above the median in Extraversion.

Students Tending Toward Introversion

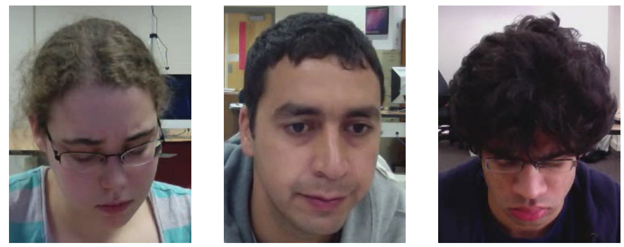
Engagement =	R^2	p
-0.6217 * AU7 $\rightarrow$ AU4	0.1302	< 0.001
-0.3999 * AU14 $\rightarrow$ CompileError	0.2285	0.005
0.3356 * MoveForward $\rightarrow$ TaskPause	0.1383	0.004
0.3246 * TypingAck $\rightarrow$ AU4	0.0764	0.020
0.3434 (Intercept)		1.000
Leave-One-Out Cross-Validated $R^2 = 0.5734$		

Table 7: Predictive model for standardized engagement in students scoring below the median in Extraversion.



AU1 $\rightarrow$ Explain (Tutor) | AU4 $\rightarrow$ AU1 | Info (Tutor) $\rightarrow$ AU1

(a) Features included in the predictive model for engagement in extraverts.



AU7 $\rightarrow$ AU4 | AU14 $\rightarrow$ C. Error | TypingAck $\rightarrow$ AU4

(b) Features included in the predictive model for engagement in introverts.

Figure 4: Selected features of the predictive models built on student engagement.

associated with posture; half of the selected features involved a movement backward in the seat. This movement was always predictive of a decrease in frustration. It is possible that this movement is part of the realization of a 'power pose', in which a person alters his posture in order to appear more dominant [9]; however, further advanced postural tracking methods would be necessary to perceive this level of nuanced movement.

Introverts, on the other hand, may indicate their own internal state more often with postural movement. Both of the predictive models built for introverts included features involving postural shifts. Learning gain was negatively associated with backward postural movement, perhaps indicating discomfort or anxiety in introverts. In the predictive model for engagement, however, a forward movement was correlated with increased engagement. Previous studies have discovered that extraverts shift in their seat more often [2]. Considering that introverts are naturally more reserved individuals [19], the fact that postural movement appears as significant in the predictive models for introverted students suggests that the movement of introverts may be more meaningful than that of extraverts.

7. CONCLUSION

The field of multimodal learning analytics is the focus of growing interest in recent years. Rich multimodal analysis of natural language and behavioral communication holds the potential to inspire great strides in developing truly adaptive, context-sensitive learning systems. Whereas prior work has examined methods for automatically identifying incoming user personality traits, the present analysis has expanded upon this line of study by taking the next step of identifying differences in behavior that contribute to the outcomes of the interaction. In this study, the outcomes considered were student learning gain at the end of the tutorial session, self-reported frustration at the end of the session, and student response to a User Engagement Survey. The results identified a variety of personality-specific behaviors that may contribute or detract from each of these goals; for example, we find that extraverts are more sensitive to tutor dialogue than introverts, and the models highlight the importance of detecting posture, particularly for introverts. There are several promising directions for future work. First, we must consider how tutors, whether human or automated, should respond to student multimodal behaviors. We should also continue to build rich models of how tutor or system events are responded to by the student, either verbally or nonverbally, in order to develop more personalized student models. The study presented in this paper has provided evidence that students of differing personalities may indicate the same mental states with different interaction patterns. Harnessing these differences to inform intelligent agents and user models will be crucial to the development of the next generation of truly adaptive systems.

Acknowledgements

The authors wish to thank the members of the LearnDialogue group 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 and IIS-1409639, and the STARS Alliance, CNS-1042468. 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.

8. REFERENCES

[1] O. Aran and D. Gatica-Perez. Cross-domain personality prediction: from video blogs to small group meetings. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, pages 127–130, Sydney, Australia, 2013. ACM.

[2] O. Aran and D. Gatica-Perez. One of a Kind: Inferring Personality Impressions in Meetings. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, pages 11–18, Sydney, Australia, 2013. ACM.

[3] R. S. J. d. Baker, S. K. D'Mello, M. T. Rodrigo, and A. C. Graesser. Better to Be Frustrated Than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-affective States During Interactions with Three Different Computer-based Learning Environments. *Int. J. Hum.-Comput. Stud.*, 68(4):223–241, Apr. 2010.

[4] J.-I. Biel and D. Gatica-Perez. The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs. *IEEE Transactions on Multimedia*, 15(1):41–55, Jan. 2013.

[5] J.-I. Biel, L. Teijeiro-Mosquera, and D. Gatica-Perez. FaceTube: Predicting Personality from Facial Expressions of Emotion in Online Conversational Video. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, pages 53–56, Santa Monica, California, USA, 2012. ACM.

[6] J.-I. Biel, V. Tsiminaki, J. Dines, and D. Gatica-Perez. Hi YouTube!: Personality Impressions and Verbal Content in Social Video. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, pages 119–126, Sydney, Australia, 2013. ACM.

[7] B. S. Bloom. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6):4–16, 1984.

[8] R. A. Calvo and S. K. D'Mello. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 2010.

[9] D. R. Carney, A. J. Cuddy, and A. J. Yap. Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance. *Psychological Science*, 21(10):1363–1368, Oct. 2010.

[10] P. A. Cohen, J. A. Kulik, and C.-L. C. Kulik. Educational Outcomes of Tutoring: A Meta-analysis of Findings. *American Educational Research Journal*, 19(2):237–248, 1982.

[11] L. R. Goldberg. The structure of phenotypic personality traits. *American Psychologist*, 48(1):26–34, 1993.

[12] S. D. Gosling, P. J. Rentfrow, and W. B. Swann. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6):504–528, Dec. 2003.

[13] 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 ACM International Conference on Multimodal Interaction (ICMI)*, pages 145–152, Santa Monica, California, USA, 2012. ACM.

[14] 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 International Conference on Educational Data Mining (EDM)*, pages 43–50, Memphis, Tennessee, USA, 2013.

[15] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester. Automatically Recognizing Facial Indicators of Frustration: A Learning-Centric Analysis. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 159–165, Geneva, Switzerland, 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 *To appear in Proceedings of the 7th International Conference on Educational Data Mining*, London, United Kingdom, 2014.

[17] 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. Number July, pages 247–256, 2012.

[18] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52:139–183, 1988.

[19] 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.

[20] 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.

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

[22] J. D. Marx and K. Cummings. Normalized change. *American Journal of Physics*, 75(1):87, 2007.

[23] 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 (IJLT)*, 8(4):382–403, 2013.

[24] 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*, 61(1):50–69, 2010.

[25] S. Oviatt and A. Cohen. Written and Multimodal Representations As Predictors of Expertise and Problem-solving Success in Mathematics. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, pages 599–606, Sydney, Australia, 2013. ACM.

[26] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang. Human Computing and Machine Understanding of Human Behavior: A Survey. In *Artifical Intelligence for Human Computing*, pages 47–71. Springer Berlin Heidelberg, New York, New York, USA, 2007.

[27] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media*. Cambridge University Press, 1997.

[28] S. Scherer, N. Weibel, L.-P. Morency, and S. Oviatt. Multimodal Prediction of Expertise and Leadership in Learning Groups. In *Proceedings of the 1st International Workshop on Multimodal Learning Analytics*, pages 1–8, Santa Monica, California, USA, 2012. ACM Press.

[29] M. R. Symonds and A. Moussalli. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, 65(1):13–21, Aug. 2010.

[30] A. Tapus and M. J. Mataric. Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance. In *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*, pages 133–140, 2008.

[31] A. K. Vail and K. E. Boyer. Adapting to Personality Over Time: Examining the Effectiveness of Dialogue Policy Progressions in Task-Oriented Interaction. In *To appear in Proceedings of the 15th Annual SIGDIAL Meeting on Discourse and Dialogue*, Philadelphia, Pennsylvania, USA, 2014.

[32] 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*, Honolulu, Hawaii, USA, 2014.