Analyzing Learner Affect in a Scenario-Based Intelligent Tutoring System

Benjamin Nye[1(✉)], Shamya Karumbaiah[2], S. Tugba Tokel[3], Mark G. Core[1], Giota Stratou[1], Daniel Auerbach[1], and Kallirroi Georgila[1]

[1] Institute for Creative Technologies, University of Southern California,
Los Angeles, USA
{nye,core,stratou,auerbach,kgeorgila}@ict.usc.edu
[2] College of Information and Computer Sciences,
University of Massachusetts Amherst, Amherst, USA
shamya@cs.umass.edu
[3] Department of Computer Education and Instructional Technology,
METU, Ankara, Turkey
stugba@metu.edu.tr

Abstract. Scenario-based tutoring systems influence affective states due to two distinct mechanisms during learning: (1) reactions to performance feedback and (2) responses to the scenario context or events. To explore the role of affect and engagement, a scenario-based ITS was instrumented to support unobtrusive facial affect detection. Results from a sample of university students showed relatively few traditional academic affective states such as confusion or frustration, even at decision points and after poor performance (e.g., incorrect responses). This may show evidence of "over-flow," with a high level of engagement and interest but insufficient confusion/disequilibrium for optimal learning.

1 Introduction

Emotions and affective reactions provide insight into the processes of academic cognition, perceptions, and mental events that cannot be directly measured. A growing amount of literature has studied academic emotions during computer-based learning, with affect measured using techniques such as self-report, human observation, text analysis, facial cues, speech analysis, physical sensors (pressure, conductance), and inferences from patterns of learner task behavior. Within the space of learning environments that have been studied, some consensus has emerged about the utility of four key cognitive-affective states: engagement/flow, confusion/disequilibrium, frustration, and disengagement/boredom [2].

Scenario-based intelligent tutoring systems (ITS), such as role-playing and simulations, have unique issues that make them more complex with respect to academic emotions. First, tutoring behavior (e.g., feedback) is often distinct from the reactions and consequences that occur during the scenario itself. Second, scenario-based learning is more likely to have a continuous assessment space

E. André et al. (Eds.): AIED 2017, LNAI 10331, pp. 544–547, 2017.
DOI: 10.1007/978-3-319-61425-0_60

(e.g., partial credit). Scenario-based tutors can also cause real or perceived time-pressure, such as ongoing system dynamics (e.g., flight simulators) or expectations (e.g., conversational norms). These issues result in a trade-off between balancing the sense of immersion in the scenario against breaking flow to encourage reflection on one's actions.

Affect has not been extensively studied in scenario-based ITS. Research on the Crystal Island ITS studied affect through behavioral patterns (e.g., time/interaction-based engagement) and building self-report into in-scenario interactions [6,7]. Replicating prior work, engagement was associated with better learning outcomes [7]. However, this methodology was limited in that it did not allow continuous moment-to-moment measures of multiple facets of affect. This is important, because it is not well-established that these four affective states operate identically during scenarios as compared to more abstract learning tasks. For example, emotions that might be considered analogous to engagement (e.g.,"invigoration") have sometimes shown the opposite of expected effects, and been associated with higher cognitive load and worse retention of skills on later tests for skills such as medical interventions [3].

The goal of this work was to observe the relationship of emotions to other components of the experience (e.g., correctness of answers, student traits) in a scenario-based ITS. Overall, while certain results replicate insights from prior work (e.g., confusion preceding incorrect answers), as a whole this research indicates that scenario-based tutoring may display different patterns of affect than observed in a traditional ITS.

2 Data Collection and Methodology

Data was collected on learners using the Emergent Leader Immersive Training Environment (ELITE)-Lite system, which was instrumented to collect a corpus of 30–60 min video logs of student interaction via laptop web cameras. ELITE is a scenario-based ITS which uses multiple-choice-based role-playing interactions to train basic counseling skills and practice them with a virtual human, while a virtual coach pro-actively provides hints and feedback [4]. An overview of experimental procedure, overall learning gains, the impact of hints, and student traits has been previously reported [1]. Data for 39 participants at a private university in California was collected across two randomly-assigned conditions with one condition always giving hints/feedback for mixed answers (Always-Mixed-Guidance) and the other never giving hints/feedback for mixed (No-Mixed-Guidance). Of these 39 participants (10 female), the majority of participants identified as Asian/Pacific Islander (33).

We use the acronym C-CERT to refer to commercial video analysis software based on the Computer Expression Recognition Toolbox [5], which performs real time facial expression recognition. C-CERT processed the video logs outputting evidence levels for 20 facial action units (AUs) and prototypical expressions such as Confusion, Frustration, Sadness, Joy, Anger, and Fear. A sweep examining various block sizes (1s to 5s) was briefly explored and 3s found to be the most interpretable. Observed rates of emotions such as frustration, and overt inattention were so rare that C-CERT data was reduced to a more limited set of categories, including only Baseline, Confusion, and Other, where Other was calculated as the maximum of all other remaining C-CERT emotions.

3 Results and Analysis

Means of the emotion categories showed that there were high evidence levels for Baseline (0.92, SD = 0.52), relatively low levels of Other emotions (0.25, SD = 0.44) and the mean for Confusion was negative (−0.7, SD = 0.68) suggesting absence. As would be expected, Pearson's correlations showed significant pairwise correlations between each emotion overall ($p < 0.01$ for all). Baseline was negatively correlated with Confusion ($r = -.35$) and Other ($r = -.88$), with Confusion positively correlated with Other ($r = .36$).

Under Pearson's correlations adjusted for repeated measures, few self-reported traits showed statistically significant results for this sample size, with only Experience and Anxiety (e.g., test anxiety) notable. Experience was positively correlated with Baseline ($r = .40$, $p < .01$) and negatively with Other ($r = -.40, p < .01$). Lack of anxiety was negatively correlated with Confusion ($r = -.34$, $p < .05$), with students who reported more academic anxiety also showing more confusion.

Compared to overall affect, affect around responses showed lower levels of Baseline (0.64, SD = 0.57) and higher Confusion (−0.41, SD = 0.70), with Other remaining similar (0.20, SD = 0.44). Pearson's correlations showed significant pairwise correlations ($p < .00$) that followed the same trends as the overall affect: Baseline was negatively correlated with Confusion ($r = -.40$) and Other ($r = -.72$), with Confusion positively correlated with Other ($r = .47$). Both overall and around responses, Baseline and Other showed no effect by condition (i.e., no difference despite more/less hint support). However, around student responses, Confusion showed statistically significant differences, with more Confusion in the No-Mixed-Guidance condition with fewer hints ($F(1, 37) = 6.25$, $p < .01$).

To look at this further, a paired t-test was conducted for each emotion covering 3s Before and 3s After of Correct, Incorrect, and Mixed answers. Confusion was significantly higher Before Incorrect versus Before Correct ($t(39) = 3.75$, $p < .00$) and also higher After Incorrect versus After Correct ($t(39) = 3.67$, $p < .00$). Confusion was also higher when comparing Incorrect and Mixed for Before ($t(39) = 3.75$, $p < .00$) and After ($t(39) = 3.48$, $p < .00$). Conversely, Other was higher Before Correct than Before Incorrect ($t(39) = 2.31$, $p < .03$). In general, this confirms that participants who showed confusion were less likely to respond correctly. However, analyses found that answer correctness was a weak predictor of After Confusion ($R^2 = 0.1$; 10-fold subject-level CV). This finding was similar for Baseline ($R^2 = 0.1$) and Other ($R^2 = 0.06$).

Regression analysis results, examining the extent to which Correctness of the answers could be predicted from Before Confusion, Before Other, Before Baseline, Time Taken, and Whether the Question had been seen before, indicated a statistically significant model (R^2 = .05, F (5,1953) = 20.13, p < .00). By comparison, adding a parameter for Question Difficulty raised fit to $R^2 = 0.18$ with 88% of variance explained loaded on Difficulty. As such, regression models indicate that the added predictive value of Confusion for Correctness may be limited.

4 Discussion and Conclusions

Overall, learners in the ELITE-Lite scenario-based ITS showed relatively low levels of affect and a Baseline facial state was dominant. Given the available data, we cannot exclude the possibility that the particular subject population showed particularly flat affect. However, while that may play a role, we believe that the primary cause for the limited incidence of academic emotions was due to sense of flow in the scenario. Considering the low levels of confusion, near-absence of frustration, and no signs of overt disengagement, evidence indicates that learners were in an engaged/equilibrium state as per D'Mello and Graesser's model [2]. This state might be thought of as "over-flow," where learners are engaged in the experience and content, but float past their failures and potential impasses.

Acknowledgment. The effort described here is sponsored by the U.S. Army Research Laboratory (ARL) under contract number W911NF-14-D-0005. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

1. Core, M.G., Georgila, K., Nye, B.D., Auerbach, D., Liu, Z.F., DiNinni, R.: Learning, adaptive support, student traits, and engagement in scenario-based learning. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) (2016)
2. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learn. Instr. **22**(2), 145–157 (2012)
3. Fraser, K., Ma, I., Teteris, E., Baxter, H., Wright, B., McLaughlin, K.: Emotion, cognitive load and learning outcomes during simulation training. Med. Educ. **46**(11), 1055–1062 (2012)
4. Hays, M.J., Campbell, J.C., Trimmer, M.A., Poore, J.C., Webb, A.K., King, T.K.: Can role-play with virtual humans teach interpersonal skills? In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) (2012)
5. Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., Bartlett, M.: The computer expression recognition toolbox (CERT). In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 298–305 (2011)
6. Robison, J., McQuiggan, S., Lester, J.: Evaluating the consequences of affective feedback in intelligent tutoring systems. In: Affective Computing and Intelligent Interaction ACII 2009, pp. 1–6. IEEE (2009)
7. Rowe, J.P., Shores, L.R., Mott, B.W., Lester, J.C.: Integrating learning, problem solving, and engagement in narrative-centered learning environments. Int. J. Artif. Intell. Educ. **21**(1–2), 115–133 (2011)