Emotional, Epistemic, and Neutral Feedback in AutoTutor Trialogues to Improve Reading Comprehension

Shi Feng, Janay Stewart, Danielle Clewley, and Arthur C. Graesser

University of Memphis
Institute for Intelligent Systems
Memphis, TN, USA.
{sfeng, jstwart7, dnclwley,a-graesser}@memphis.edu

Abstract. We manipulated three types of short feedback (emotional, epistemic, and neutral) in an intelligent tutoring system designed to help struggling adult readers improve reading comprehension strategies. We conducted our research on college students to eventually compare with the targeted adult population. We also recorded their facial emotions. Although participants self-reported a preference for emotional feedback, there were no differences in individual motivation or usefulness ratings between emotional and epistemic feedback. Analysis from coded facial emotions indicated that participants tended to be more sensitive to epistemic feedback than emotional feedback when using AutoTutor-CSAL.

Keywords: feedback · intelligent tutoring systems· emotions · agents

1 Introduction

Feedback has been widely shown to be an important facilitator of learning and performance [1]. Recently, an ITS, called AutoTutor-CSAL (Center for the Study of Adult Literacy [2]), has been developed that has a teacher agent (Cristina) and a student agent (Jordan) trialogue system to assist struggling adult readers. Feedback is implemented in this system as an important feature to enhance learning gains and motivation.

In the framework of AutoTutor-CSAL, feedback is usually a short general emotive feedback response to immediately let the learner know his/her performance on a single item [3], such as "_name_, you were right! Great effort!" and its variant. The learner also receives an explanation (called "answer") for the correct answer after receiving a short emotive feedback. These short feedback speech acts are designed to mimic feedback given by human tutors, which would increase the learner's engagement [4]. One open research question is whether giving short emotive feedback that emphasizes positive or negative emotions in a feedback benefits both the learner's engagement and performance, compared with an epistemic feedback which simply informs the learner whether his/her input on any given item is correct or incorrect, or

adfa, p. 1, 2011.

Feng, S., Stewart, J., Clewley, D., & Graesser, A. C.(2015). Emotional, epistemic, and neutral feedback in AutoTutor trialogues to improve reading comprehension. In C. Conati, N. Heffernan, A. Mitrovic, M. F. Verdejo (Eds.), *Proceedings of the 17th International Conference on Artificial Intelligence in Education* (pp. 570-573). Cham: Springer.

a neutral feedback that simply acknowledges the learner's input. We know that most human tutors tend to be reluctant to give negative feedback when the learner responds incorrectly [5], perhaps due to fear of lowering the learner's confidence.

The aim of the present research is to investigate whether different types of short feedback would affect learners' engagement and performance in a usability study with college students interacting with AutoTutor CSAL. Specifically, we examined whether short emotional feedback, commonly utilized in lessons in AutoTutor CSAL, would improve or engage learners' performance better than short epistemic or neutral feedback. If emotional feedback motivates learners in a learning environment, then there should be an increase in response time, since in the context of reading, it was previously found that time spent on reading has a strong relationship with motivation [6], higher performance, and a better learner experience using the system. Furthermore, we analyzed the participants' facial emotions while interacting with AutoTutor CSAL. Past research has found six emotions that frequently occur during learning. Those emotions are: confusion, boredom, frustration, flow, delight, and surprise [7]. Students vary in experiencing these emotions during learning. By tracking participants' moment-to-moment affective states, we can get a better sense of their user-experience when they encounter different types of feedback after participants answer comprehension questions.

2 Method

Participants were 63 college undergraduates from the University of Memphis Subject Pool who participated in this study for course credit. The experiment used a within-subjects design with three feedback conditions: short emotive feedback ("_user_, you were correct. Good job!"/ "unfortunately, _user_, you did not get it correct this time"), short epistemic feedback ("_user_, you were correct"/ "_user_, you were incorrect"), and short neutral feedback ("ok"). The learning materials consisted of a lesson from AutoTutor CSAL designed to teach learners how to comprehend narrative texts using the reading strategy called summarization. The college participants read all three texts of easy, medium and difficult levels of 4.5, 5.9, and 7.7 Flesch-Kincaid Grade Levels [8]. The easy text was a 700-word short story version of Puccini's opera, *Turandot*. The medium text was an 852-word short story version of Eminescu's poem, *The Legend of the Evening Star*. The difficult text was a 760-word short story version of Shakespeare's play, *The Twelfth Night*. All three short stories were written and adapted from their original source by a graduate student. After reading each text, participants were presented with comprehension questions, each tailored to a specific portion of the text. After giving the feedback, Cristina always gave an elaborate explanation for the correct answer choice. Order of feedback condition, and order of texts with the comprehension questions that come after each text, were counterbalanced across participants with a Graeco-Latin Square to specify the assignment of the texts to treatment conditions. During the course of learning, their facial expressions were recorded with the Morae recorder, which recorded both participants' facial expressions. After each interaction with each text, participants were asked to fill out a questionnaire to rate their impressions and experiences using the system on a 1 to 5

Likert Scale, with 1 representing strongly disagree with a positive aspect of the feedback condition and 5 representing strongly agreeing with a positive aspect of the feedback condition. The scale of 3 indicated that the feedback feature is not important to the learning environment.

3 Results and Discussion

There was no significant difference between the average correct responses on a comprehension question item, either between feedback conditions or between text conditions (M = .86, SD = .34). We investigated whether the response time between the feedback conditions varies between the different texts to which they were answering the comprehension questions. We found that there was a significant main effect of response time between the three texts conditions (F(2, 180)=5.35, p <.01). The interaction was significant (F(4,180) = 3.77, p = .003). Only the medium difficulty level text ("*The Legend of the Evening Star*") showed a significant mean difference of response time between the three feedback conditions (F(2, 690) = 5.123, p = .006). According to a Tukey Post Hoc test, while reading the medium difficulty level text, participants spent significantly longer time answering the comprehension questions when they received emotional feedback. This suggests that they may be more motivated when they received emotional feedback while answering questions from the medium text. (See Table 1).

Table 1. *Mean response times (seconds) to comprehension questions per text condition as a function of feedback conditions.*

	Neutral Feedback	Epistemic Feedback	Emotional Feedback
Text 1 (Easy)	6.63 (13.80)	5.36 (9.28)	4.66 (8.32)
Text 2 (Medium)	7.75 (12.37)	7.32 (10.29)	11.18 (18.57)*
Text 3 (Difficult)	6.49 (12.28)	8.86 (16.61)	7.81 (14.51)

* *Significant response time with respect to Alpha = .05; SD in parenthesis*.

We were particularly interested in the participants' change of affective state when they received the first positive feedback versus the first negative feedback statement within each feedback condition. Three students coded the participants' facial affective states on the six learner-centered emotions in the places where the participants had the first correct and the first incorrect response within 2 seconds. Each participant was rated on a 1 to 6 signal detect scale segregating absence (1-3) and present of emotion (4-6) for each of the six emotions. The rating of confidence for each side of the segregation varies so that the low number signifies the lowest confidence in ratings. Cronbach's Alpha for continuous scale showed high inter-rater reliability of .86 on the items. The three independent ratings were then averaged to produce a mean score for each emotion per participant. The data are presented in Table 2. Overall, participants are significantly more likely to be frustrated when they received the first negative feedback statement regardless of feedback condition (emotional: t(41) = 2.11, p = .044; epistemic: t(36) = 3.16, p = .002; neutral: t(43) = 2.13, p = .039). They were

more likely to be confused when they received the first negative emotional feedback ($t(41) = 2.04$, $p = .047$). They were more engaged (in flow state) when they received the first positive epistemic feedback ($t(36 = 2.93$, $p = .008$), and more bored and surprised when they received the first negative epistemic feedback (boredom: $t(36) = 2.61$, $p = .014$; surprise: $t(36) = -2.61$, $p = .014$). These results show that participants in general are most sensitive to epistemic feedback in terms of affective state change.

Table 2. *Mean ratings of learning emotions for the first positive and first negative feedback as a function of feedback conditions*

	Emotional Feedback		**Epistemic Feedback**		**Neutral Feedback**	
	Pos. Feed	Neg. Feed	Pos. Feed	Neg. Feed	Pos. Feed	Neg. Feed
Boredom	2.95 (1.20)	3.67 (1.23)	2.43 (.80)*	3.20 (.98)*	2.75 (1.21)	3.11 (1.14)
Flow	4.35 (.87)	3.95 (.99)	4.81 (.51)**	4.22 (.70)**	4.67 (.89)	4.19 (.97)
Frustration	1.38 (.41)*	1.85 (.95)*	1.43 (.54)**	2.16 (.82)**	1.25 (.46)*	1.63 (.70)*
Confusion	1.35 (.53)*	1.77 (.80)*	1.62 (.74)	1.94 (.66)	1.33 (.46)	1.61 (.54)
Delight	1.14 (.36)	1.09 (.23)	1.32 (.63)	1.24 (.54)	1.11 (.29)	1.03 (.09)
Surprise	1.16 (.66)	1.38 (.74)	1.12 (.22)*	1.50 (.56)*	1.16 (.54)	1.28 (.59)

*Notes: *significant mean difference between positive and negative feedback with respect to alpha = .05; **significant at alpha = .01; SD in parenthesis.*

The majority (66%) of the participants reported that they liked the emotional feedback the best. Four mixed-effect linear modeling approach was adopted to analyze the four impression ratings of usefulness ("Do you agree that this feedback is useful in this learning environment?"), naturalness ("Do you find it natural to give this type of feedback?"), confidence boosting ("Do you find your confidence being boosted when receiving this type of feedback?), and motivation ("do you find it motivating to receive this type of feedback?). The unit of analysis was individual impression ratings (207 cases). The random effect was participants (63 levels). Feedback condition was a three-level categorical fixed effect, and ratings were a continuous dependent variable. All four models were significant (usefulness: $\eta^2 = 0.04$, $F(2, 204) = 12.06$, $p < .001$; naturalness: $\eta^2 = 0.01$, $F(2, 203) = 4.59$, $p = .011$; confidence boosting: $\eta^2 = 0.03$, $F(2, 204) = 6.60$, $p = .002$; motivation: $\eta^2 = 0.05$, $F(2, 203) = 12.10$, $p < .001$). Tukey Post Hoc tests showed that emotional and epistemic did not significantly differ in these ratings, but both were rated higher than neutral.

4 General Discussion

Although participants reported preferring emotional feedback over epistemic or neutral feedback, they overall rated epistemic feedback to be just as useful and motivating as emotional feedback. Therefore it is not clear that giving emotional short feedback has benefits over epistemic in terms of engaging the learner in our ITS with AutoTutor agents. Our results suggest that the effect of giving emotional feedback in order to engage the learner on a task (reading comprehension) may vary with the con-

tent and nature of the materials. Our participants spent the majority of their time in the affective state of flow or engagement, which is in accordance to previous findings [7], but our research also revealed that participants' facial emotion tended to be the most sensitive to the epistemic feedback. For example, the positive emotion of delight was generally rated higher when participants received epistemic feedback than neutral or emotional, although our methods of determining delight may overlap with frustration to some degree [9]. Similarly, the emotion of surprise, which usually occurs when a learner receives a negative feedback [10], also seems to occur more frequently when participants receive epistemic feedback. One explanation is that the nature of the task was overall a bit too easy for the college population. The short emotional feedback had little benefit perhaps because it appeared patronizing or ineffectual, whereas an epistemic feedback is more motivating by virtue of its being succinct and clear. Nevertheless, the question remains whether struggling adult readers will show a different profile of responses to feedback.

Acknowledgments. This research was supported by the Institute of Education Sciences (IES) (R305C120001).

References

1. Shute, V.J.: Focus on Formative Feedback. Review of Educational Research. 78(1), 153–189 (2008).
2. Graesser, A.C., Baer, W., Feng, S., Walker, B., Clewley, D., Hayes, D.P.: Emotions in Adaptive Computer Technologies for Adults Improving Reading. In: Tettegah, S., Ferdig, R. (eds.) Emotions and Technology: Communication of Feeling For, With, and Through Digital Media. pp. 1--35. Elsevier, San Diego (2014).
3. Cai, Z., Feng, S., Baer, W., Graesser, A.C.: Instructional Strategies in Trialoguebased Intelligent Tutoring Systems. In: Goldberg, B., Graesser, A.C., Hu, X., and Sottilare, R.: (eds.) Design Recommendations for Intelligent Tutoring Systems. vol. 2, pp. 225-235. U.S. Army Research Laboratory, Orlando, FL (2014).
4. Price, M., Handley, K., Millar, J., O'Donovan, B.: Feedback: All that Effort, but What is the Effect?. Assessment & Evaluation in Higher Education. 35(3), 277–89 (2010).
5. Person, N.K., Kreuz, R.J., Zwaan, R., Graesser, A.C.: Pragmatics and Pedagogy: Conversational Rules and Politeness Strategies May Inhibit Effective Tutoring. Cognition and Instruction. 13, 161–188 (1995).
6. Guthrie, J.T., Wigfield, A., Metsala, J.L. Cox, K.E.: Motivational and Cognitive Predictors of Text Comprehension and Reading Amount. Scientific Studies of Reading. 3, 231-256 (1999).
7. D'Mello, S.K.: A Selective Meta-analysis on the Relative Incidence of Discrete Affective States During Learning with Technology. Journal of Educational Psychology. 105, 1082--1099 (2013).
8. Klare, G.R.: Assessing Readability. Reading Research Quarterly. 10, 62--102 (1974).
9. Hoque, M.E., McDuff, D.J., Picard, R.W: Exploring Temporal Patterns in Classifying Frustrated and Delighted Smiles. Affective Computer. 3, 323-334 (2012).
10. D'Mello, S.K., Graesser, A.C.: AutoTutor and Affective AutoTutor: Learning by Talking with Cognitively and Emotionally Intelligent Computers that Talk Back. ACM Transactions on Interactive Intelligent Systems. 2(4), 1--38 (2012).