


3-28-2018

Affect-Aware Adaptive Tutoring Based on Human–Automation Etiquette Strategies

Euijung Yang
Iowa State University

Michael C. Dorneich
Iowa State University, dorneich@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/imse_pubs

 Part of the [Ergonomics Commons](#), [Industrial and Organizational Psychology Commons](#), [Industrial Engineering Commons](#), and the [Systems Engineering Commons](#)

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/imse_pubs/179. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Industrial and Manufacturing Systems Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Industrial and Manufacturing Systems Engineering Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Affect-Aware Adaptive Tutoring Based on Human–Automation Etiquette Strategies

Abstract

Objective: We investigated adapting the interaction style of intelligent tutoring system (ITS) feedback based on human–automation etiquette strategies.

Background: Most ITSs adapt the content difficulty level, adapt the feedback timing, or provide extra content when they detect cognitive or affective decrements. Our previous work demonstrated that changing the interaction style via different feedback etiquette strategies has differential effects on students' motivation, confidence, satisfaction, and performance. The best etiquette strategy was also determined by user frustration.

Method: Based on these findings, a rule set was developed that systemically selected the proper etiquette strategy to address one of four learning factors (motivation, confidence, satisfaction, and performance) under two different levels of user frustration. We explored whether etiquette strategy selection based on this rule set (systematic) or random changes in etiquette strategy for a given level of frustration affected the four learning factors. Participants solved mathematics problems under different frustration conditions with feedback that adapted dynamic changes in etiquette strategies either systematically or randomly.

Results: The results demonstrated that feedback with etiquette strategies chosen systematically via the rule set could selectively target and improve motivation, confidence, satisfaction, and performance more than changing etiquette strategies randomly. The systematic adaptation was effective no matter the level of frustration for the participant.

Conclusion: If computer tutors can vary the interaction style to effectively mitigate negative emotions, then ITS designers would have one more mechanism in which to design affect-aware adaptations that provide the proper responses in situations where human emotions affect the ability to learn.

Keywords

adaptive automation, affective factors, intelligent tutors, human–computer interaction, etiquette

Disciplines

Ergonomics | Industrial and Organizational Psychology | Industrial Engineering | Systems Engineering

Comments

This is a manuscript of an article published as Yang, Euijung, and Michael C. Dorneich. "Affect-Aware Adaptive Tutoring Based on Human–Automation Etiquette Strategies." *Human Factors* (2018). DOI: [10.1177/0018720818765266](https://doi.org/10.1177/0018720818765266). Posted with permission.

Affect-Aware Adaptive Tutoring Based on Human-Automation Etiquette Strategies

Euijung Yang, Michael C. Dorneich¹

Iowa State University, Ames, IA, USA

Abstract

Objective. This study investigated adapting the interaction style of intelligent tutoring system (ITS) feedback based on human-automation etiquette strategies.

Background. Most ITSs adapt the content difficulty level, feedback timing, or provide extra content when it detects cognitive or affective decrements. Our previous work demonstrated that changing the interaction style via different feedback etiquette strategies have differential effects on students' motivation, confidence, satisfaction, and performance. The best etiquette strategy was also determined by user frustration.

Method. Based on these findings, a rule set was developed that systemically selected the proper etiquette strategy to address one of four learning factors (motivation, confidence, satisfaction, and performance) under two different levels of user frustration. An experiment explored whether etiquette strategy selection based on this rule set (systematic) or random changes in etiquette strategy for a given level of frustration impacted the four learning factors. Participants solved mathematics problems under different frustration conditions with feedback that adapted dynamic changes in etiquette strategies either systematically or randomly.

Results. The results demonstrated that feedback with etiquette strategies chosen systematically via the rule set could selectively target and improve motivation, confidence, satisfaction, and performance more than changing etiquette strategies randomly. The systematic adaptation was effective no matter the level of frustration for the participant.

Conclusion. If computer tutors can vary the interaction style to effectively mitigate negative emotions, then ITS designers would have one more mechanism in which to design affect-aware adaptations that provide the proper responses in situations where human emotions affect the ability to learn.

Keywords: Adaptive Automation, Affective Factors, Intelligent Tutors, Human-Computer Interaction, Etiquette

¹ Correspondence concerning this article should be addressed to Michael C. Dorneich, Iowa State University, 3004 Black Engineering Building, Ames, IA 50011-2164. Contact: dorneich@iastate.edu

INTRODUCTION

Cognition and emotion (affect) are complementary processes that enable student learning (Lehman, D'Mello, & Person, 2010). Positive emotions can result in improved information processing decision making, creative problem solving, cognitive flexibility, and memory retention (Contai & Manske, 2009; Erez & Isen, 2002; Isen, Daubman, & Nowicki, 1987; Isen, Shalker, Clark, & Karp, 1978). Negative emotions, notably frustration, can lower productivity, increase decision-making time, and lower learning efficiency (Klein, Moon, & Picard, 2002; Powers, Rauh, Henning, Buck, & West, 2011; Lerner, Li, Valdesolo, & Kassam, 2015; Graesser, Chipman, Haynes, & Olney, 2005).

Tutors are keenly aware of the importance of emotions, and actively use strategies to influence human emotions to improve student learning (Petrovica, 2013). Tutors may avoid disagreement to maintain non-adversarial relationships, or remain optimistic to encourage students when they struggle (Pearson, Kreuz, Zwaan, & Graesser, 1995). Tutors spend as much time attending to student emotional needs as cognitive learning needs (Lepper & Chabay, 1988). Intelligent tutoring systems (ITSs) are computer-based instructional systems that provide one-to-one feedback to enable learning by modifying instructional content, timing, and teaching strategies (Wenger, 1987; Murray, 2003; Koedinger & Tanner, 2013; Gilbert, Blessing, & Guo, 2015). ITS design has more frequently focused on cognitive aspects of learning, such as assessing student content knowledge to trigger tutor feedback (Wood & Wood, 1999; Zakharov, Mitrovic, & Johnston, 2008; Roll, Alevan, McLaren, & Koedinger, 2011). This contrasts with human tutors, who adapt behavior to meet the emotional needs of students (Woolf et al., 2009).

Evaluating learning effectiveness using only assessments of performance ignores other factors influencing learning. Keller (2009) identified multiple learning factors for sustaining student motivation and performance: Attention, Relevance, Confidence, and Satisfaction (ARCS). Higher levels of these factors correlate with higher levels of engagement with the material (Mohammad & Job, 2012) and increased learning effectiveness (Keller, 2009). ITSs have not always achieved the same impact as a good human tutor (Fletcher-Flinn & Gravatt, 1995). The ability to incorporate affect into ITS systems may partially address this gap (Zakharov et al., 2008). Researchers have developed affect-aware systems with the ability to detect and adapt to user emotional states (Picard, 1997; D'Mello et al., 2008; Woolf et al., 2009; Wang & Johnson, 2008; Grawemeyer et al., 2015).

Much work has focused on the detection of emotion. For instance, emotion has been shown to be related to mental workload and mental engagement, two constructs that can be detected through EEG-based indexes (Chaouachi, & Frasson, 2012). One barrier to wide application has been the difficulty in accurately classifying emotional states in a low-cost, unobtrusive manner.

Increasingly researchers are also studying how to prevent negative emotional states (Zakharov et al., 2008), regulate emotions (Malekzadeh, Mustafa, & Lahsasna, 2015), and mitigate the effects of emotions (Petrovica, 2013; Jraidi, Chalfoun, & Frasson, 2012; D'Mello, & Calvo, 2011). While some emotions (delight, boredom) are susceptible to the appropriateness of the feedback, negative emotions such as frustration have been more robust, where any attempt to alleviate frustration usually results in some benefit to the student (Robison, McQuiggan, & Lester, 2009). Theoretical frameworks can inform intervention design. For instance, self-regulation models (Boekaerts & Corno, 2005) address the ability of students to mitigate emotions to stay in the zone of proximal development; neither too frustrated or too bored (Murray & Arroyo, 2002). Negative emotions can decrease self-regulation, leading to poorer learning outcomes, with the opposite effect for positive emotion (Pekrun, Goetz, Titz, & Perry, 2002). Recent work has focused on the role confusion plays in benefitting learning (D'Mello, Lehman, Pekrun, & Graesser, 2014). Confusion during the learning process can be beneficial if students work to overcome the source of misunderstandings, if they remain engaged and motivated (Robison, McQuiggen, and Lester, 2009).

Adaptive ITSs automatically tailor behavior in real-time to best support learning. Adaptive systems have four general categories of modification: function allocation, task scheduling, interaction style, and content (Feigh, Dorneich, & Hayes, 2012). Adaptations to negative emotions have typically been task-focused or emotion-focused actions (Malekzadeh et al., 2015). Task-related strategies include changing task difficulty (Harley, Lajoie, Frasson, & Hall, 2015), adjusting timing and difficulty of assessments (Arroyo et al., 2014), or providing additional examples and hints (Chaffar, Derbali, & Frasson, 2009; Woolfe et al, 2009). Emotion-focused actions include providing empathetic responses (D'Mello, & Graesser, 2012; Mao & Li, 2009), mirroring the learner emotions (Picard et al., 2004; Zakharov et al, 2008), or providing behavioral prompts (D'Mello, & Graesser, 2015).

What has been less explored is modification of the interaction. In this work, changes to the *interaction style* of an ITS are produced by changes in the politeness level of a feedback

statement, but not its amount of information or subject matter. Our work is built upon the premise, established by previous work, that intelligent tutoring systems should employ social intelligence, where *how* a tutor provides feedback is as important as the cognitive intelligence of *when* and *what* feedback is presented (McLaren, DeLeeuw, & Mayer, 2011; Lepper & Woolverton, 2002). This work explores the viability of changing the *interaction style* of ITSs when providing feedback, without changing content difficulty level, feedback timing, or feedback information content. The concept of etiquette strategies may provide design inspiration on how to adapt interaction style.

Social interactions between humans are governed by expectations based on conventional norms. Etiquette is a code of requirements for social behaviors (Brown & Levinson, 1978). The linguistic aspects of etiquette of human-human interaction have identified independent factors in politeness including social power (e.g., relative positions in social hierarchies, age, gender), social distance (i.e., politeness increases with familiars, but decreases with both intimates and strangers), and imposition (e.g., requesting, urgency, apologizing, thanking, indebtedness, complaining). For a review, see Kasper (2005). Social power and distance are decided by the relationship between speaker and hearer. Level of imposition refers to the amount of demand or burden on the hearer. While the social power and the social distance between two people only change slowly over time, the imposition from speaker to hearer can be adjusted moment-to-moment.

Brown and Levinson (1987) identified four types of etiquette strategies to protect an individual's self-esteem ("face threat") from imposition: bald, negative politeness, positive politeness, and off-record (for examples, see Table 1). In bald strategy, the speaker speaks in a direct way without consideration of the imposition level on the hearer, or minimizing threats to the hearer's face. Positive politeness minimizes social distance between speaker and hearer by expressing statements of friendship, solidarity, and compliments. Negative politeness attempts to be respectful and assumes imposition on the hearer. Off-record gives indirect feedback, requiring the hearer to infer what was intended. Adaptation of the style of interaction using different etiquette strategies would alter the level of face threat which the imposition causes in communication. Pearson et al. (1995) demonstrated that human tutors varied between three different etiquette strategies when interacting with students to enhance or inhibit tutoring: bald, positive politeness, negative politeness.

Table 1. Example sentences of etiquette strategies for feedback regarding formulas (example 1) and drawing (example 2).

Etiquette Strategies	Example 1 sentences (subject is formulas)	Example 2 sentences (subject is drawing)
Bald	Use appropriate formula.	Draw the problem out to understand it better.
Positive Politeness	Why don't you try other formulas? Let's check them together!	Let's try to draw a picture to better understand what the problem is asking.
Negative Politeness	If it's alright with you, could you please check other formulas as well?	I don't mean to interrupt you but it would help if you draw out the problem.
Off-Record	Various formulas are provided.	I wonder what this problem looks like visually.

This work investigates the premise that human-human etiquette strategies are generalizable to human-computer etiquette. Users tend to personify their computers by attributing human-like characteristics to them, and react to computers based on expectations from human-human interaction (Nass, Steuer, & Tauber, 1994). The concept of human-human etiquette has been applied in human-computer interaction (Miller, Wu, & Funk, 2008; Hayes & Miller, 2010; Miller, Wu, & Ott, 2012; Dorneich, Ververs, Mathan, Whitlow, & Hayes, 2012). Studies have operationalized politeness into various types of etiquette strategies. Johnson and Rizzo (2004) developed a model of politeness in tutor dialog that incorporated politeness factors of power and distance, and learner states of motivation and confidence, to automatically generate natural language dialogs for use in a foreign language tutor that generated text-based feedback in one of multiple politeness strategies: exaggerate, common ground, be vague, understate, question, tautology, impersonalize, and indirect. McLaren et al (2007) operationalized politeness by rephrasing all feedback to be more “polite” and simultaneously improving both its positive and negative face. In the current study, we operationalized Brown and Levinson’s (1987) etiquette strategies and focused on changes to the redress of imposition of feedback.

In preliminary work, different etiquette strategies and levels of frustration combined to have different impacts on the learning factors of performance, motivation, confidence, and satisfaction (Yang & Dorneich, 2016). For instance, bald etiquette strategy led to higher performance, positive politeness to higher subjective ratings of motivation and satisfaction, and negative politeness to higher subjective ratings of confidence and satisfaction. However, the most effective strategy for a given learning factor sometimes changed with frustration. The

current study builds upon these results to develop a rule set (see next section) that specifies under which conditions (triggers) different strategies should be used to provide feedback during students' problem-solving. It is an open question whether a generalized set of rules can be effective for a population of users, and whether they can be designed in such a way to provide one-on-one interaction customized to the individual. Computer tutors more responsive to their students emotions may help students persist in their learning even when frustrated or motivation might otherwise flag.

Studies comparing polite feedback with direct feedback have mixed results. Wang and Johnson (2008) found evidence of learning gains for some types of questions but not others, and gains in some measures of motivation. In evaluations of web-based tutoring, McLaren et al. (2011) found lower prior-knowledge students benefited more from polite than direct feedback, although the overall effect was weaker in the classroom than the lab. In this current study, our aim was to test if differences within polite feedback can be used to target and improve the learning factors of motivation, confidence, satisfaction, and performance while mitigating the effect of frustration.

ADAPTIVE TUTORING SYSTEM PROTOTYPE

An adaptive tutoring prototype interface developed for this evaluation provided college-level algebra, geometry, trigonometry, calculus, statistics, and probability problems (Figure 1). Problems could be labeled as "easy" or "hard" to describe their level of difficulty. Research has shown that the level of frustration people experience can be affected by how confident they are in their ability to achieve the goal (Bessiere, Ceaparu, Lazar, Robinson, & Shneiderman, 2004). When a learner expects a problem to be easy but then finds it difficult, a learner's expectations of goal attainment and satisfaction are unfulfilled, which can lead to frustration (Berkowitz, 1989). Based on the results of a pilot study, labeling hard question as easy was used in the evaluation to manipulate frustration. In addition, the evaluation verified that this manipulation did significantly change perceived frustration in participants.

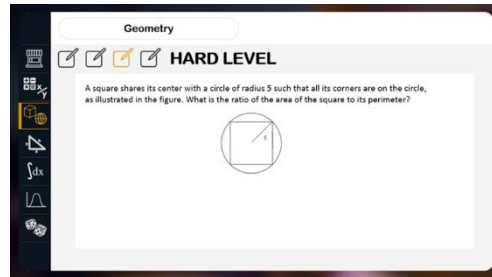


Figure 1. The screenshot of the adaptive tutoring system prototype

The system provided recorded voice feedback while students solved problems. Voice and appearance can give cues to factors such as gender, age, education level, culture, and socio-economic position. Therefore recordings were designed to be homogenous and neutral, using the same voice, intonation, accent, and speaking pace. In a Wizard-of-Oz evaluation (Dahlbäck, Jönsson & Ahrenberg, 1993), while the participant solved math problems, the experimenter observed progress and triggered appropriate feedback voice files, based on a rubric (Table 2) for specific problem-solving errors or misconceptions (Gordon, 2008).

Table 2. Errors or Misconceptions that Trigger Feedback for different problem-solving steps.

Step	Problem Solving Step	Errors or Misconceptions that Trigger Feedback
1	Identify and define the problem or situation	<ul style="list-style-type: none"> • Does not know how to start • Does not know which variables are defined • Does not know how to define variables
2	Generate alternatives	<ul style="list-style-type: none"> • Does not know which variables to solve for • Cannot find the equation to use • Does not know how to approach • Miss other variables
3	Evaluate the alternative suggestions	<ul style="list-style-type: none"> • Is not sure this approach is correct • Does not solve all variables
4	Make the decision	<ul style="list-style-type: none"> • Cannot decide on the calculation methods
5	Implement the solution or decision	<ul style="list-style-type: none"> • Does not realize they are finished

Each step had at least one proactive (e.g., “Define the variables”) and at least one reactive (e.g., “It’s not the appropriate formula”) feedback comment. Feedback was specific to each math problem and differed between problems. Four versions of each feedback, one in each etiquette strategy, were designed to be homogeneous in content by restricting each feedback utterance to one subject (e.g., formula, drawing – Table 1), and varying the politeness vocabulary. Therefore,

length of utterances differed. Future work could explore if feedback length has an impact beyond the type of politeness used.

The adaptive tutoring system was designed to improve four learning factors associated with the ARCS model: motivation, confidence, satisfaction, and performance. A rule set was developed to trigger the most appropriate etiquette strategy as the basis for systematic adaptation, by targeting the learning factor most in need of improvement and applying an etiquette strategy to improve it. The process is outlined in Figure 2.

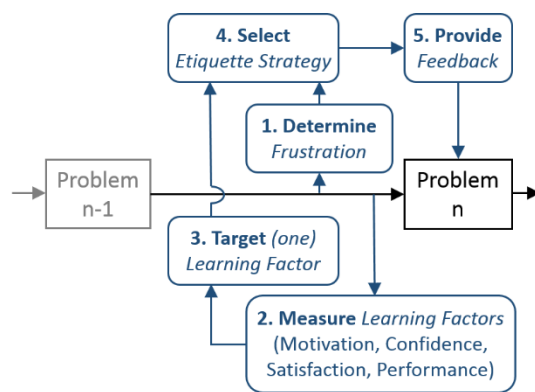


Figure 2. The logical flow of the tutor to decide which learning factor to target, and which etiquette strategy to use when providing feedback in the next problem.

After each math problem, the student rated motivation, confidence, and satisfaction on a 10-point scale, and a performance score was calculated (step 2). An algorithm determined which learning factor should be “targeted” next (step 3). First, if any learning factors’ ratings ≤ 5 , target factor with the lowest score. If multiple factors shared the same lowest rating, the ARCS model determined priority order (highest to lowest): motivation, confidence, satisfaction, performance. If all ratings > 5 , choose factor with the largest decrease from the previous measurement. Ties were broken using the same order as above.

Once the targeted learning factor (step 3) and the level of frustration (step 1) were identified, the etiquette strategy was chosen (step 4) from options listed in Table 3. Previous work established an initial, course-grained guidance on the most effective etiquette strategy given a learning factor and frustration level (Yang & Dorneich, 2016). Feedback during the next problem was delivered using the chosen etiquette strategy (step 5). Some combinations had more than one etiquette strategy identified as effective. If a strategy failed to improve the targeted

factor, then an alternative strategy was chosen the next time that factor was targeted. This enables some individual customization as the system learns which strategy is most effective.

Table 3. Etiquette strategies that improved the learning factor given a level of frustration (Yang & Dorneich, 2016).

Learning Factor	Low Frustration	High Frustration
Motivation	Positive Politeness	Positive Politeness
Confidence	1) Bald, 2) Positive Politeness, 3) Negative Politeness	1) Positive Politeness, 2) Negative Politeness
Satisfaction	Positive Politeness	1) Positive Politeness, 2) Negative Politeness
Performance	1) Bald, 2) Positive Politeness	Negative Politeness

METHOD

Independent Variables

The two independent variables were *Frustration* (low, high) and *Adaptation* (systematic, random). Frustration was manipulated by misaligning expectations of problem difficulty with reality, based on the theory that frustration comes from unfulfilled expectations (Berkowitz, 1989). Conversely, frustration may be diminished when people have lower level of expectation about their goal achievement (Bessiere et al., 2004). Recognizing a difference between the expected and actual level of difficulty can cause frustration (Hone, 2006; Glass, McGuinness, & Wolverton, 2008). All problems had a similar difficulty level, with historical Graduate Record Exam (GRE) correct answer rate of 30% – 40%. However, half of the problems were labeled as ‘easy’ even though they were just as difficult as problems labeled “hard”. In pilot tests, math problems of different difficulty levels and labels were given to participants, who self-reported their level of frustration on a 10-point scale. The final level of difficulty was chosen such that it produced enough frustration (when mislabeled as “easy”) to impact learning but not too high to cause the participant to give up.

Adaptation was manipulated by adjusting how the etiquette strategy for the next math problem was chosen: systematically or randomly. Systematic adaptation used the rule set described in the previous section. Random adaptation randomly chose etiquette strategies regardless of previous results. Systematic adaptation was compared to random adaptation rather than a no-feedback baseline since any feedback, no matter how poorly designed, could benefit

the learner due to the added information of the feedback as well as the possibility that changes in the feedback itself may cause the learner to engage with the material more. Previous work has established that feedback in the optimal etiquette strategy was significantly better than no-feedback (Yang & Dorneich, 2016). A random number generator set the transitions between etiquette strategies within the random trials. The strategy changed between each problem. It should be noted that in the random condition, there was a 25% chance that the etiquette strategy chosen for a math problem is the same one that would be chosen under the systematic condition. This may make differences harder to detect.

Hypothesis

- H1: Applying etiquette strategies based on the systematic rule set increases motivation, confidence, satisfaction, and performance more than randomly applying etiquette strategies.
- H2: Applying etiquette strategies based on the systematic rule set mitigates frustration more than random etiquette strategies.

Participants

A power analysis of the first 10 subjects indicated a minimum sample size of 21.75 to detect a significant effect for performance (power .90, alpha .05). Thirty-three university students (19 males, 14 females) participated in the experiment (demographics in Table 4), had normal or corrected-to-normal vision, and majored in STEM fields.

Table 4. Participant demographics, computer usage, and self-reported math skill ratings on a scale of 1-10.

Descriptor	Mean	Variability
Age	25.5	range: 19 – 31
Computer usage	8.0 hours/day	range: 3 – 16
algebra	8.2	<i>SD</i> = 1.5
geometry	7.0	<i>SD</i> = 1.9
trigonometry	6.8	<i>SD</i> = 1.8
calculus	7.8	<i>SD</i> = 1.8
statistics	6.6	<i>SD</i> = 2.5
probability	6.9	<i>SD</i> = 2.2
Last attended math class	2.6 years	Range: 1 – 5

Experimental Design

The experiment was a 2 (*Frustration*: low, high) x 2 (*Adaptation*: systematic, random), repeated measures, within-subject design, with one trial for each combination of independent

variables. Each trial contained five math problems, where the etiquette strategy varied either systematically or randomly between math problems, and the problems were all labeled either “hard” or “easy”. Trial order was counterbalanced using 4x4 Latin squares.

Dependent Variables

The dependent variables are described in Table 5.

Table 5. The measurements for both independent variables verification and dependent variables.

Variables	Metric	Measurement (Unit)	Frequency	Data Type
IV Manipulation Check	TLX Subscale Frustration	Scale 0 – 10	After each math problem within trial	Subjective
Task Performance	Problem Solving Score	Score 0 – 10	After each math problem within trial	Objective
Motivation	Motivation Questionnaire	Scale 0 – 10	After each math problem within trial	Subjective
Confidence	Confidence Questionnaire	Scale 0 – 10	After each math problem within trial	Subjective
Satisfaction	Satisfaction Questionnaire	Scale 0 – 10	After each math problem within trial	Subjective
Effectiveness when Targeting an Improvement in a Learning Factor	Change in level of learning factors from last math problem to the current: Δ Motivation, Δ Confidence, Δ Satisfaction, Δ Performance	(-10) – (+10)	After each math problem within trial	Objective
Feedback Effectiveness	Effectiveness Questionnaire	Scale 0 – 10	After each math problem within trial	Subjective
Feedback Appropriateness	Appropriateness Questionnaire	Scale 0 – 10	After each math problem within trial	Subjective
Cognitive Workload	TLX Subscale Mental Demand	Scale 0 – 10	After each math problem within trial	Subjective

Independent Variable Manipulation Check. To confirm that the labeling of problems induced differing levels of frustration, participants rated their frustration on the NASA TLX (Hart & Staveland, 1988) Frustration subscale after each trial.

Motivation, Confidence, and Satisfaction. After each math problem, participants rated, on a 10-point scale:

- How motivated do you feel to continue working on tasks?
- How much satisfaction did you experience based on system's feedback?
- How confident do you feel about your performance during the task?

Task Performance. The subject solved the math problem presented by the system using paper and pencil. Performance was graded (Table 6) immediately after each problem by the same experimenter who triggered feedback during the problem. The experimenter was the same for all subjects.

Table 6. The rubric to score problems.

Score	Answer Sheet
10	Answer is correct with variables and equations demonstrated
7.5	Answer has correct equation but with calculation mistakes
5.0	Answer has correct solution approach such as setting up the variable from problems or drawing shapes based on given problems, but used wrong equations or no equations
2.5	If the participant tried to make variables or draw shapes but they were not the correct approach
0	If the answer sheet does not have anything correct

Effectiveness when Targeting an Improvement in a Learning Factor. The effectiveness and tradeoffs of the mitigation on targeted and non-targeted factors was measured by the difference in ratings between two consecutive problems for the four learning factors (Δ Motivation, Δ Confidence, Δ Satisfaction, Δ Performance).

Feedback Appropriateness, Effectiveness, Cognitive Workload. Cognitive workload was measured with the mental demand subscale of the NASA TLX. Two other constructs were rated on a 10-point scale:

- Was feedback from the system appropriate?
- Was feedback from system effective to solve the task?

Procedure

This research complied with the APA Code of Ethics and was approved by the Institutional Review Board at Iowa State University (16-004). After informed consent, participants were provided training and practice to solve math problems until comfortable.

During trials, the experimenter observed the participant and triggered feedback based on the rubric (Table 2). This ensured all feedback was triggered correctly, and effects uncovered were solely due to etiquette strategies and not confounded with triggering the wrong feedback. After every math problem, participants rated motivation, confidence, and satisfaction. The experimenter calculated their performance score. The experimenter entered the ratings and performance score into a rule-set calculator (excel file) to determine the etiquette strategy for next math problem. This procedure took less than one minute. The etiquette strategy only changed between math problems. After every trial (5 math problems) participants rated workload, feedback effectiveness, and feedback appropriateness. A post-experiment survey gathered formative feedback. The debriefing revealed that the true goal was to study etiquette strategies, not mathematics ability.

Data Analysis

The Shapiro-Wilk test found that normality assumptions were not met for most metrics, so non-parametric Wilcoxon tests were used. To answer the question, “*Was the targeted learning factor improved by the system?*” a Wilcoxon analysis determined the effect of adaptation and frustration on the targeted learning factor. To answer the question, “*What also happened to the non-targeted learning factors?*” a nonparametric Steel-Dwass test was conducted on each pairwise combination of learning factors for each adaptation type. Wilcoxon results are reported as highly significant for a significance level $\alpha < .001$, significant for $\alpha < .05$, and marginally significant for $\alpha < .10$ (Gelman, 2013). In addition, since Wilcoxon tests cannot calculate an interaction effect between two factors, the Aligned Rank Transform (ART) method was used to analyze interaction effects in nonparametric factorial data (Wobbrock, Findlater, Gergle, & Higgins, 2011). ART aligned and ranked data before using regular F-tests. Cohen’s d measured effect size of the mean difference between two groups in standard deviation units, and were reported as small ($.20 < d < .50$), medium ($.50 < d < .80$), and large ($d > .80$). A Pearson product-moment correlation coefficient r_s was calculated to understand the association between

demographic data (age, skill level, years since last math class) on the learning factors. None of the demographic factors has any correlation ($|r_s| > 0.3$) to motivation, confidence, satisfaction, or performance scores.

RESULTS

Independent Variable Manipulation Check

Frustration rating was significantly ($Z = 13.6, p < .001, d = 1.09$) higher for the high frustration (“easy” label) condition than the low frustration (“hard” label) condition (Figure 3). The main effect of adaptation was not significant, $Z = 0.41, p = .68$. This verifies that the manipulation of frustration was effective.

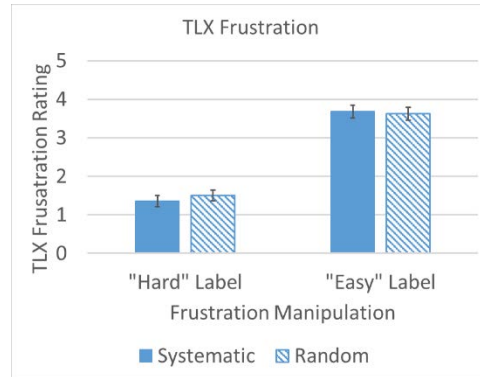


Figure 3. Mean and standard error of TLX frustration rating.

Effectiveness when Targeting an Improvement in Learning Factor

Given that any learning factor could be targeted for a math problem, the number of data points for each factor may be different.

Targeting Motivation. Figure 4 illustrates the change in motivation (Δ Motivation) when targeting motivation. The main effect of adaptation was significant, $Z = 2.14, p = .033, d = 0.50$. Systematic adaptation resulted in a larger Δ Motivation ($M = 2.8, SD = 4.2, N = 77$) than random adaptation ($M = 1.0, SD = 3.1, N = 62$). The main effect of frustration on Δ Motivation was not significant, $Z = 0.30, p = .76$. The interaction effect was non-significant, $F(1,16) = 0.66, p = .43$.

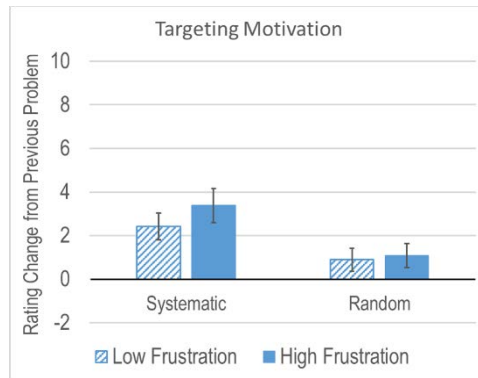


Figure 4. Mean and standard error of motivation rating change.

Figure 5 illustrates how learning factors were affected when motivation was targeted. Since the main effect of frustration was not significant, the data were collapsed across this variable. For the systematic adaptation, Figure 5 indicates significant pairwise differences between groups when they do not share a capital letter. Δ Motivation was significantly larger than the other three learning factors, with a large average effect size of $d = 0.97$. For the random adaptation, Figure 5 indicates significant pairwise differences between groups when they do not share a lower-case letter. Δ Motivation was significantly larger than Δ Satisfaction and Δ Performance, with a small average effect size of $d = 0.45$.

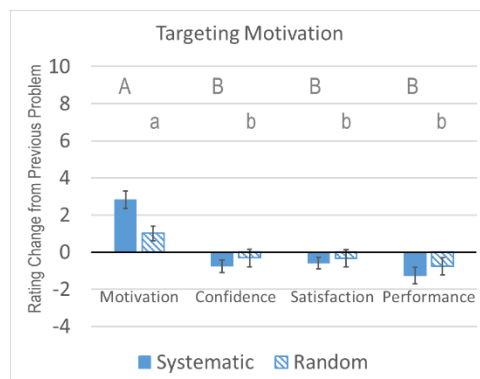


Figure 5. Mean and standard error of Δ Motivation with systematic ($n=77$) and random adaptation ($n=62$) for both the targeted (motivation) and non-targeted learning factors. A significant pairwise difference between groups is indicated when they do not share a capital letter (Systematic) or lower case letter (Random).

Targeting Confidence. The main effect of adaptation was significant, $Z = 2.72$, $p = .006$, $d = 0.56$ (Figure 6). Systematic adaptation resulted in a larger Δ Confidence ($M = 4.3$, $SD = 3.52$, $N = 60$) than random adaptation ($M = 2.0$, $SD = 4.5$, $N = 58$). The main effect of frustration on Δ Confidence was not significant, $Z = 0.78$, $p = .44$. The interaction effect was non-significant, $F(1,19) = 0.41$, $p = .53$.

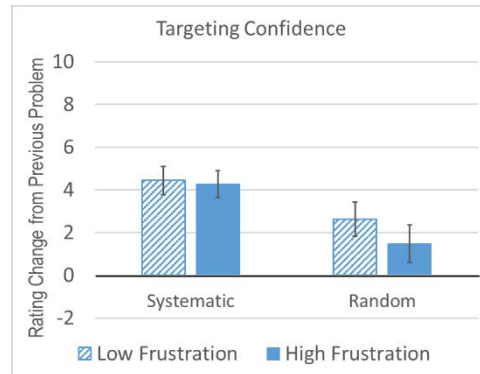


Figure 6. Mean and standard error of confidence rating change.

For systematic adaptation, Δ Confidence was significantly larger than the three non-targeted learning factors, with a large average effect size of $d = 1.09$ (Figure 7). For random adaptation, Δ Confidence was significantly larger than three non-targeted learning factors, with a moderate average effect size of $d = 0.58$.

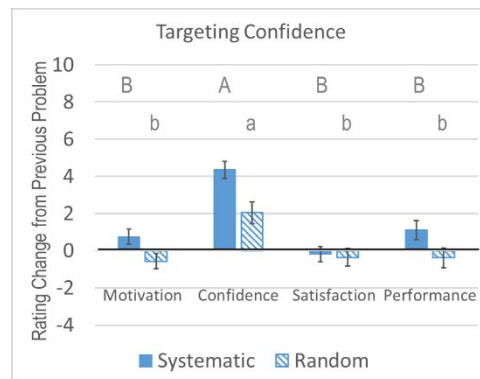


Figure 7. Mean and standard error of motivation with systematic ($n=60$) and random adaptation ($n=58$) for both the targeted (confidence) and non-targeted learning factors. A significant pairwise difference between groups is indicated when they do not share a capital letter (Systematic) or lower case letter (Random).

Targeting Satisfaction. The main effect of adaptation was marginally significant, $Z = 1.89$, $p = .058$, $d = 0.42$ (Figure 8). Systematic adaptation resulted in a larger Δ Satisfaction ($M = 4.0$, $SD = 3.9$, $N = 60$) than random adaptation ($M = 2.3$, $SD = 4.1$, $N = 59$). The main effect of frustration on Δ Satisfaction was not significant, $Z = 0.07$, $p = .94$. The interaction effect was non-significant, $F(1,10) = 0.11$, $p = .75$.

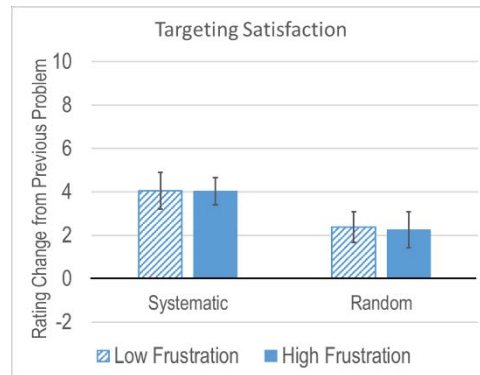


Figure 8. Mean and standard error of satisfaction rating change.

For systematic adaptation, Δ Satisfaction was significantly larger than the three non-targeted learning factors, with a large average effect size of $d = 1.03$ (Figure 9). For random adaptation, Δ Satisfaction was significantly larger than three non-targeted learning factors, with a moderate average effect size of $d = 0.58$.

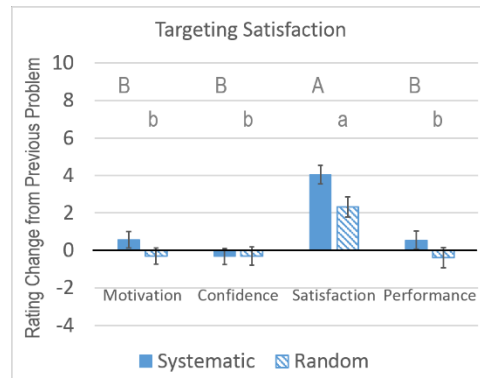


Figure 9. Mean and standard error of motivation with systematic ($n=60$) and random adaptation ($n=59$) for both the targeted (Satisfaction) and non-targeted learning factors. A significant pairwise difference between groups is indicated when they do not share a capital letter (Systematic) or lower case letter (Random)

Targeting Performance. The main effect of adaptation was significant, $Z = 2.50$, $p = .012$, $d = 0.47$ (Figure 10). Systematic adaptation resulted in a larger Δ Performance ($M = 2.5$, $SD = 3.6$, $N = 62$) than random adaptation ($M = 0.73$, $SD = 3.6$, $N = 82$). The main effect of frustration on

Δ Performance was not significant, $Z = 0.32, p = .75$. The interaction effect was non-significant, $F(1,24) = 2.76, p = .11$.

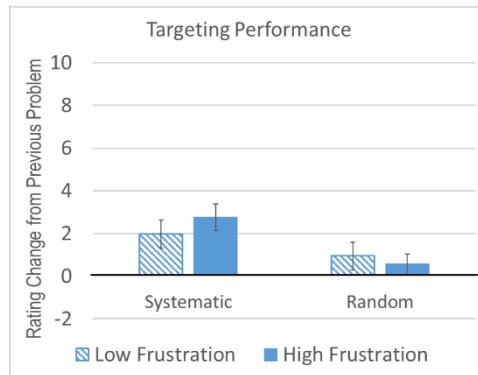


Figure 10. Mean and standard error of performance rating change.

For systematic adaptation, Δ Performance was significantly larger than the three non-targeted learning factors, with a large average effect size of $d = 0.87$ (Figure 11). For random adaptation, Δ Performance was significantly larger than Δ Confidence and Δ Satisfaction, with a small average effect size of $d = 0.41$.



Figure 11. Mean and standard error of motivation with systematic (n=62) and random adaptation (n=82) for both the targeted (Performance) and non-targeted learning factors. A significant pairwise difference between groups is indicated when they do not share a capital letter (Systematic) or lower case letter (Random)

Feedback Effectiveness

The main effect of adaptation on feedback effectiveness was not significant, $Z = 0.88, p = .38$ (Figure 12). The main effect of frustration on feedback effectiveness was marginally significant, $Z = 1.80, p = .072, d = .21$. The effectiveness rating was higher in the high frustration condition ($M = 5.8, SD = 3.7$) than in the low frustration condition ($M = 5.3, SD = 3.7$). The

interaction effect was non-significant, $F(1,29) = 0.15, p = .70$. There was no significant effect of targeted learning factor on feedback effectiveness, $Z = 3.53, p = .31$.

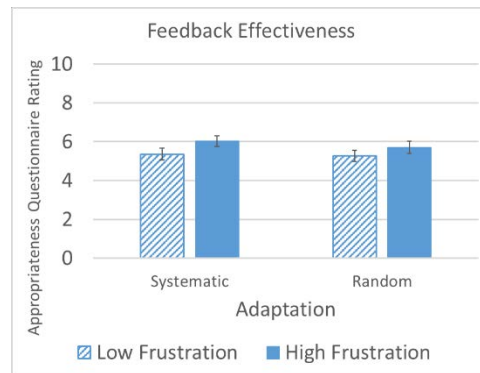


Figure 12. Mean and standard error of feedback effectiveness.

Feedback Appropriateness

The main effect of adaptation on feedback appropriateness was not significant, $Z = 0.88, p = .38$ (Figure 13). The main effect of frustration on feedback appropriateness was marginally significant, $Z = 1.80, p = .072, d = .19$. The appropriateness rating was higher in the high frustration condition ($M = 6.3, SD = 3.5$) than in the low frustration condition ($M = 5.8, SD = 3.6$). The interaction effect was non-significant, $F(1,30) = 1.81, p = .19$. There was no significant effect of targeted learning factor on feedback appropriateness, $Z = 1.32, p = .72$.

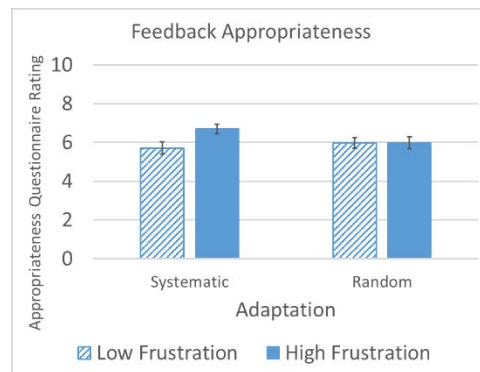


Figure 13. Mean and standard error of feedback appropriateness.

Cognitive Workload

The main effect of adaptation on the TLX mental demand rating was not significant, $Z = 0.78$, $p = .44$ (Figure 14). The main effect of frustration on mental demand was significant, $Z = 3.36$, $p < .001$, $d = 0.49$. Mental demand was higher in the high frustration condition ($M = 3.2$, $SD = 2.2$) than in the low frustration condition ($M = 2.4$, $SD = 4.5$). The interaction effect was non-significant, $F(1,33) = 1.01$, $p = .32$.

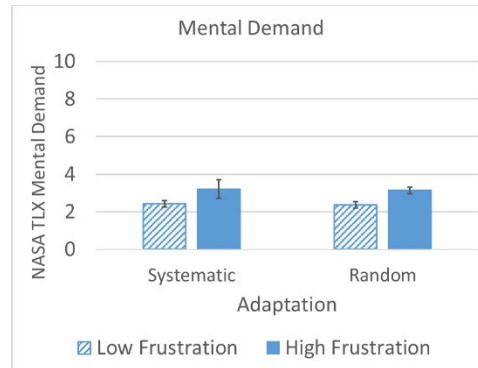


Figure 14. Mean and standard error of TLX mental demand.

DISCUSSION

Systematically adapting interaction style based on etiquette strategies significantly influenced motivation, confidence, satisfaction, and performance. Hypothesis H1 was fully supported. The targeted learning factor under systematic adaptation increased significantly more than under random adaptation. Hypothesis H2 was partially supported. Systematic adaptation significantly improved targeted learning factors in both high and low frustration conditions, with the same level of improvement regardless of frustration level. However, frustration did not affect the random condition significantly either. This aligns with the observation of Robison et al (2009) that negative emotions such as frustration may benefit from interventions even when inappropriately delivered. However, this could also be because the high frustration condition averaged a rating of four out of 10, possibly a level not high enough to significantly affect the learning process. Although even at this moderate level of frustration, there was a marginally significant, albeit small effect on the appropriateness and effectiveness ratings of the feedback.

Frustration may affect performance (Libb, 1972) because it indicates dissatisfaction related with encountered difficulties in learning (Radford, 2015). Frustration induced in this work was one of a mismatch of expectations to actual difficulties. Future work will study the effect of

higher levels of frustration on the learning factors. Future work can also explore other sources of frustration including task difficulty itself (Rosenzweig, 1938), time pressure (Wahlström, Hagberg, Johnson, Svensson, & Rempel, 2002), ineffective or irrelevant feedback (McKinney, 1933; Baylor & Rosenberg-Kima, 2006), or human-computer interaction issues (Klein et al., 2002; Powers et al., 2011).

The study also investigated potential adaptation tradeoffs. Systematic adaptation improved targeted learning factors significantly more than non-targeted learning factors. An ITS system would have to decide when a small negative change in the non-targeted factor was acceptable for the larger gains in the targeted factor. Previous work had established that “polite” feedback was more effective than “direct” feedback for some measures of motivation and performance (Wang et al., 2005; McLaren et al., 2011). This work demonstrates that “polite” feedback can systematically be adapted during instruction to individually target and improve motivation, confidence, satisfaction, and performance, where systemic application outperforms random (but still polite) application.

However, the random adaptation also significantly improved the targeted learning factor compared to non-targeted factors. But where effect sizes for systematic adaptation were large, effect sizes in the random adaptation were small. There are several possibilities why there was an improvement in random adaptation at all. Post hoc analyses determined that 21.2% of the time the random choice of etiquette strategy matched the result that would have been obtained by using the systematic rules. A second possibility is that the activation caused by any change at all resulted in improvement. However, none of the non-targeted learning factors showed improvements and so this possibility is less likely. Finally, there might be an adjustment towards the mean, where the learning factor most in need of improvement (the targeted factor) would improve the most in the next trial, no matter what etiquette strategy was applied. If this was the case then the random and systematic conditions would see similar levels of improvement in the targeted factor. However, the magnitude of the improvement under systematic adaptation was 2.5x larger (range: 1.7 – 3.4) compared to random adaptation.

A good human tutor is aware of learner emotional state and adapts their interaction style to support factors that underlie performance such as a motivation, confidence, or satisfaction. However, it can be difficult for a human tutor to determine how best to respond to a student’s emotional needs, depending on the human tutor’s ability to both detect the emotion correctly, and

then have a strategy on how to respond (Petrovica, 2013). There is a lack of a theoretical framework to determine how an ITS can use emotions in learning situations (Petrovica, 2013). Developing affect-aware ITS are still in early stages, and much more work is needed to understand the full impact of the use of emotions in tutoring systems (D'Mello & Calvo, 2012). Leveraging previous work in human-human and human-computer etiquette, this work aims to provide mechanisms to enable an ITS to adapt its interaction style to impact the learning factors that support performance as identified by Keller (2009). The evaluation results demonstrate that not only do changing etiquette strategies have a differential effect on student learning factors, but that those differential effects can be used systemically to target and improve learning factors. However, many research questions remain. Changes in etiquette strategy happened between each problem, but human tutors may be able to change their strategies during a problem. Thus, future work may look at moment-to-moment detection of human affect to trigger changes in real-time. Additionally, this work focused on testing the effectiveness of the adaptations and assumed "perfect triggering" to initiate feedback by having the experimenter decide when to trigger feedback instead of sensors to detect learner state. This ensured that any effects could be attributed to the etiquette strategies. However, a fully closed-loop affect-aware ITS would automatically detect frustration, learning factors, and problem-solving difficulties in real time. Thus, two lines of research are needed to address this gap: 1) automated detection of frustration and learning factors and 2) ITS logic that incorporates imperfect detection. Using sensors to detect emotional state is an active area of research, and includes video of facial expressions (Loijens et al., 2012), facial electromyography (Hazlett, 2003), or skin conductance (Boucsein, 2012). Every detection system has some level of uncertainty (i.e. "imperfect triggering"), thus further work is needed to understand how the trustworthiness, uncertainty, and fuzziness of emotion-detection systems impact overall ITS effectiveness (Landowska, 2013). For instance, there may be an interaction between inappropriate feedback (due to an incorrect detection of emotional state) and motivation and trust. An ITS that attempts to display higher levels of emotional intelligence and support may invoke higher expectations on the part of learners. If the systems "get it wrong" then trust may erode, which may lead to lower motivation, confidence and satisfaction. Understanding what the threshold of accuracy is needed before providing the wrong interaction style is worse than making no change at all is an open question that future research must address.

Although this study looked only at modifications of interaction (*how*), fully formed adaptive ITSs would still consider modification of content (*what*) and modification of task scheduling (*when*). For instance, confidence could be targeted by presenting easy questions in a lesson first, before progressing to more difficult items. Future work will study the interplay of the timing, content, and style of adaptive feedback to improve learners cognitive and affective states. Individual differences of the participants must be considered. The ultimate goal is to develop a library of possible options (of feedback content and etiquette strategy) and a rule set that triggers the appropriate feedback to best meet the learning objectives. Finally, this experiment used only math problems. The type of task may influence which etiquette strategy is optimal under what conditions. A future area of research would be to expand this adaptive interaction style approach to non-STEM fields.

Acknowledgements

The authors would like to thank Mariangely Iglesias-Pena, David Montealegre, Jordan Zonner, and Maria Dropps for supporting experiment design and data analysis. The authors would like to thank Leslie Potter and Dr. Caroline Hayes for feedback on the manuscript. This material is based in part upon work supported by the National Science Foundation under Grant No. 1461160.

References

- Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, 24(4), 387-426.
- Baylor, A. L., & Rosenberg-Kima, R. B. (2006). Interface agents to alleviate online frustration. In *Proceedings of the 7th international conference on Learning sciences* (pp. 30-36). International Society of the Learning Sciences.
- Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, 54(2), 199-231.
- Berkowitz, L. (1989). Frustration-aggression hypothesis: examination and reformulation. *Psychological bulletin*, 106(1), 59.
- Bessiere, K., Ceaparu, I., Lazar, J., Robinson, J., and Shneiderman, B. 2004. Social and psychological influences on computer user frustration. In *Media access: Social and*

- psychological dimensions of new technology use, eds. E. P. Bucy and J. E. Newhagen, pp. 91–103. Mahwah, NJ: Lawrence Erlbaum Associates.
- Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.
- Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In *Questions and politeness: Strategies in social interaction* (pp. 56-311). Cambridge University Press.
- Chaffar, S., Derbali, L., & Frasson, C. (2009). Inducing positive emotional state in Intelligent Tutoring Systems. In *AIED* (Vol. 2009, No. 200, pp. 716-718).
- Chaouachi, M., & Frasson, C. (2012, June). Mental workload, engagement and emotions: an exploratory study for intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems* (pp. 65-71). Springer Berlin Heidelberg.
- Conati, C., & Manske, M. (2009, September). Evaluating adaptive feedback in an educational computer game. In *International Workshop on Intelligent Virtual Agents* (pp. 146-158). Springer Berlin Heidelberg.
- Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—why and how. *Knowledge-based systems*, 6(4), 258-266.
- D’Mello, S. K., & Graesser, A. C. (2015). Feeling, thinking, and computing with affect-aware learning. *The Handbook of Affective Computing* (Calvo, R.A., D’Mello, S.K., Gratch, J., Kappas, A., eds.), Oxford University Press. 419-434.
- D’Mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(4), 23.
- D’Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153-170.
- D’Mello, S. K., & Calvo, R. A. (2011). Significant accomplishments, new challenges, and new perspectives. In *New perspectives on affect and learning technologies* (pp. 255-271). Springer New York.
- D’Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R. & Graesser, A., (2008). AutoTutor detects and responds to learners affective and cognitive states. In *Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems* (pp. 306-308).

- Dorneich, M. C., Ververs, P. M., Mathan, S., Whitlow, S., & Hayes, C. C. (2012). Considering etiquette in the design of an adaptive system. *Journal of Cognitive Engineering and Decision Making*, 6(2), 243-265.
- Erez, A., & Isen, A. M. (2002). The Influence of positive affect on the components of expectancy motivation. *Journal of Applied Psychology*, 87(6), 1055–1067.
doi:10.1037/0021-9010.87.6.1055
- Feigh, K.M., Dorneich, M.C., & Hayes, C.C. (2012). “Towards a Characterization of Adaptive Systems: a Framework for Researchers and System Designers”, *Journal of Human Factors and Ergonomics*. Vol. 54, No. 6, pp. 1008-1024.
- Fletcher-Flinn, C. M., & Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): A meta-analysis. *Journal of educational computing research*, 12(3), 219-241.
- Gelman, A. (2013). Commentary: P values and statistical practice. *Epidemiology*, 24(1), 69-72.
- Gilbert, S. B., Blessing, S. B., & Guo, E. (2015). Authoring Effective Embedded Tutors: An Overview of the Extensible Problem Specific Tutor (xPST) System. *International Journal of Artificial Intelligence in Education*, 1-27.
- Glass, A., McGuinness, D. L., & Wolverton, M. (2008, January). Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces* (pp. 227-236). ACM
- Gordon, T. (2008). *Parent effectiveness training: The proven program for raising responsible children*. Harmony.
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *Education, IEEE Transactions on*, 48(4), 612-618.
- Grawemeyer, B., Mavrikis, M., Holmes, W., Gutierrez-Santos, S., Wiedmann, M., & Rummel, N. (2016, April). Affecting off-task behaviour: how affect-aware feedback can improve student learning. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 104-113). ACM.
- Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2015, June). An integrated emotion-aware framework for intelligent tutoring systems. In *International Conference on Artificial Intelligence in Education* (pp. 616-619). Springer International Publishing.

- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.
- Hayes, C. C. & Miller, C. A. (2010). Should computers be polite? In C. C. Hayes & C. A. Miller (Eds.), *Human-computer etiquette: Cultural expectations and the design implications they place on computers and technology* (pp. 1–12). Boca Raton, FL: Taylor & Francis.
- Hazlett, R. (2003). Measurement of user frustration: a biologic approach. In *CHI'03 extended abstracts on Human factors in computing systems* (pp. 734-735). ACM.
- Hone, K. (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers*, 18(2), 227-245.
- Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology*, 52(6), 1122–1131. doi:10.1037/0022-3514.52.6.1122
- Isen, A. M., Shalcker, T. E., Clark, M., & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology*, 36(1), 1–12. doi:10.1037/0022-3514.36.1.1
- Johnson, W. L., & Rizzo, P. (2004). Politeness in tutoring dialogs: “run the factory, that’s what I’d do”. In *Intelligent Tutoring Systems* (pp. 206-243). Springer Berlin/Heidelberg.
- Jraidi, I., Chalfoun, P., & Frasson, C. (2012, June). Implicit strategies for intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems* (pp. 1-10). Springer Berlin Heidelberg.
- Kasper, G. (2005). Linguistic etiquette. *Intercultural discourse and communication: The essential readings*, 58-67.
- Keller, J. M. (2009). *Motivational design for learning and performance: The ARCS model approach*. Springer Science & Business Media.
- Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with computers*, 14(2), 119-140.
- Koedinger, K., & Tanner, M. (2013). Things you should know about intelligent tutoring systems. *EDUCAUSE Learning Initiative (ELI)*.
- Landowska, A. (2013, June). Affect-awareness framework for intelligent tutoring systems. In *Human System Interaction (HSI), 2013 The 6th International Conference on* (pp. 540-547). IEEE.

- Lehman, B., D'Mello, S., & Person, N. (2010, June). The intricate dance between cognition and emotion during expert tutoring. In *International Conference on Intelligent Tutoring Systems* (pp. 1-10). Springer Berlin Heidelberg.
- Lepper, M. R., & Chabay, R. W. (1988). Socializing the intelligent tutor: Bringing empathy to computer tutors. In *Learning issues for intelligent tutoring systems* (pp. 242-257). Springer US.
- Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. *Improving academic achievement: Impact of psychological factors on education*, 135-158.
- Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Psychology*, 66.
- Libb, J. (1972). Frustration and task complexity: an extension of frustration theory. *Journal of experimental psychology*, 96(1), 67.
- Loijens, L., Krips, O., Van Kuilenbug, H., Den Uyl, M., Ivan, P., Theuws, H., and Spink, A. (2012) *FaceReader Reference Manual Version 5.0*, Noldus Information Technology b.v.
- Malekzadeh, M., Mustafa, M. B., & Lahsasna, A. (2015). A Review of Emotion Regulation in Intelligent Tutoring Systems. *Educational Technology & Society*, 18(4), 435-445.
- Mao, X., & Li, Z. (2009, April). Implementing emotion-based user-aware e-learning. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 3787-3792). ACM.
- McKinney, F. (1933). Certain emotional factors in learning and efficiency. *The Journal of General Psychology*, 9(1), 101-116.
- McLaren, B. M., Lim, S. J., Yaron, D., & Koedinger, K. R. (2007). Can a Polite Intelligent Tutoring System Lead to Improved Learning Outside of the Lab?. *Frontiers in Artificial Intelligence and Applications*, 158, 433.
- McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011). Polite web-based intelligent tutors: Can they improve learning in classrooms?. *Computers & Education*, 56(3), 574-584.
- Miller, C. A., Wu, P., & Funk, H. B. (2008). A computational approach to etiquette: Operationalizing Brown and Levinson's politeness model. *Intelligent Systems, IEEE*, 23(4), 28-35.

- Miller, C. A., Wu, P., & Ott, T. (2012). Politeness in Teams: Implications for Directive Compliance Behavior and Associated Attitudes. *Journal of Cognitive Engineering and Decision Making*, 6(2), 214-242.
- Mohammad, S., & Job, M.A. (2012). Confidence-Motivation-Satisfaction-Performance (CMSP) Analysis of Blended Learning System in the Arab Open University Bahrain. *International Journal of Information Technology and Business Management*, 3(1).
- Murray, T. (2003). An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art. In *Authoring tools for advanced technology learning environments* (pp. 491-544). Springer Netherlands.
- Murray, T., & Arroyo, I. (2002). Toward measuring and maintaining the zone of proximal development in adaptive instructional systems. In *Intelligent tutoring systems* (pp. 749-758). Springer Berlin/Heidelberg.
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. pp. 72-78.
- Pearson, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and instruction*, 13(2), 161-188.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, 37(2), 91-105.
- Petrovica, S. (2013). Adaptation of tutoring to students' emotions in emotionally intelligent tutoring systems. In *e-Learning and e-Technologies in Education (ICEEE), 2013 Second International Conference on* (pp. 131-136). IEEE
- Picard, R.W. (1997). *Affective computing*. Cambridge: MIT press.
- Picard, R.W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D. & Strohecker, C. (2004). Affective learning—a manifesto. *BT Technology Journal*, 22(4), 253-269.
- Powers, S. R., Rauh, C., Henning, R. A., Buck, R. W., & West, T. V. (2011). The effect of video feedback delay on frustration and emotion communication accuracy. *Computers in Human Behavior*, 27(5), 1651-1657.

- Radford, L. (2015). Of love, frustration, and mathematics: A cultural-historical approach to emotions in mathematics teaching and learning. In *From beliefs to dynamic affect systems in mathematics education* (pp. 25-49). Springer International Publishing.
- Robison, J., McQuiggan, S., & Lester, J. (2009, September). Evaluating the consequences of affective feedback in intelligent tutoring systems. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on* (pp. 1-6). IEEE.
- Roll, I., Alevan, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267-280.
- Rosenzweig, S. (1938). The experimental measurement of types of reaction to frustration. *Explorations in personality, 585-99*.
- Wahlström, J., Hagberg, M., Johnson, P., Svensson, J., & Rempel, D. (2002). Influence of time pressure and verbal provocation on physiological and psychological reactions during work with a computer mouse. *European journal of applied physiology, 87*(3), 257-263.
- Wang, N., & Johnson, W. L. (2008, June). The Politeness Effect in an intelligent foreign language tutoring system. In *International Conference on Intelligent Tutoring Systems* (pp. 270-280). Springer, Berlin, Heidelberg.
- Wang, N., Johnson, W. L., Rizzo, P., Shaw, E., & Mayer, R. E. (2005, January). Experimental evaluation of polite interaction tactics for pedagogical agents. In *Proceedings of the 10th international conference on Intelligent user interfaces* (pp. 12-19). ACM.
- Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems*. Los Altos, CA: Morgan Kaufmann.
- Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 143-146). ACM.
- Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education, 33*(2), 153-169.
- Woolf, B., Bursleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognizing and responding to student affect. *International Journal of Learning Technology, 4*(3-4), 129-164.

Yang, E. & Dorneich, M. C. (2016) Evaluation of Etiquette Strategies to Adapt Feedback in Affect-Aware Tutoring. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Zakharov, K., Mitrovic, A., & Johnston, L. (2008). Towards emotionally-intelligent pedagogical agents. In International Conference on Intelligent Tutoring Systems (pp. 19-28). Springer Berlin Heidelberg.

Biographies

Dr. Euijung Yang is a human factors engineer and user experience researcher at ConnectFive. She earned her PhD in human factors in the Industrial and Manufacturing Systems Engineering Department at Iowa State University in 2016. She earned her BS and MS at in Industrial Engineering at Soongsil University.

Dr. Michael C. Dorneich is an Associate Professor at Iowa State University in the Industrial and Manufacturing Systems Engineering Department since 2012. Previously he was a Principal Research Scientist in the Human Centered Systems department at Honeywell Laboratories. He earned his PhD in Industrial Engineering in the human factors program at the University of Illinois at Urbana-Champaign in 1999.