Responding to Student Uncertainty During Computer Tutoring: An Experimental Evaluation

Kate Forbes-Riley, Diane Litman, and Mihai Rotaru

University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA 15260
forbesk,litman,mrotaru@cs.pitt.edu

Abstract. This paper evaluates dialogue-based student performance in a controlled experiment using versions of a tutoring system with and without automatic adaptation to the student affective state of uncertainty. Our performance metrics include correctness, uncertainty, and learning impasse severities, which are measured in a "test" dialogue after the tutoring treatment. Although these metrics did not significantly differ across conditions when considering all student answers in our test dialogue, we found significant differences in specific types of student answers, and these differences suggest that our uncertainty adaptation does have a positive benefit on student performance.

1 Introduction

In recent years, tutoring researchers have shown increasing interest in the interplay between student affect and learning (e.g. [1,2,3]). Numerous tutoring dialogue system researchers are investigating the hypothesis that student performance can be improved by automatically detecting and adapting to affective states (e.g., [4,5,6,7]). Student uncertainty is one state of primary interest due to its theorized relationship to correctness and learning. Researchers hypothesize that uncertainty can signal to the tutor that there is an opportunity for learning to occur, and that experiencing uncertainty can motivate a student to engage in learning (e.g. [6,8,9]). Moreover, correlational studies have shown a link between uncertainty and learning (e.g. [6]). However, few controlled experiments have investigated the performance impact of uncertainty adaptations in computer tutoring; most computer tutors respond based only on student correctness.

Based on this prior research, we hypothesized that responding to uncertainty - in addition to correctness - should improve student performance. We tested this hypothesis in a controlled experiment using adaptive and non-adaptive versions of a spoken dialogue tutoring system. Uncertainty and correctness were manually annotated in real-time by a human "Wizard". The experiment had three conditions. In the experimental condition, the system provided additional knowledge at places of uncertainty. In one control condition, the system did not provide this knowledge after uncertainty; in a second control condition the system provided this knowledge randomly.

B. Woolf et al. (Eds.): ITS 2008, LNCS 5091, pp. 60–69, 2008.

Section 2 of this paper describes the experiment.[1] Section 3 presents a comparison of student performance metrics across condition. Section 4 discusses the implications of these results. Section 5 explains how we used these results to improve the design of a larger version of this experiment that is now underway.

2 The Experiment

In prior work we developed ITSPOKE (**I**ntelligent **T**utoring **SPOKE**n dialogue system) [11], a spoken dialogue tutor that is built on top of the Why2-Atlas text-based tutor [12] and tutors 5 qualitative physics problems. The spoken dialogues have a Question - Answer - Response format, implemented with a finite state dialogue manager. ITSPOKE responses (states) depend only on the correctness of the student answer (transitions between states). If the answer is correct, ITSPOKE moves on to the next question. ITSPOKE responses to incorrect answers take two forms: 1) For incorrect answers to easier questions, ITSPOKE provides the correct answer with a brief statement of reasoning. 2) For incorrect answers to harder questions, ITSPOKE engages the student in a **remediation subdialogue**, containing questions that walk the student through the more complex line of reasoning required for the correct answer.

2.1 Adaptive Wizard-of-Oz Spoken Dialogue Tutoring System

We've begun enhancing ITSPOKE to automatically respond to student affect[2] over and above correctness. For two reasons, we have initially targeted uncertainty. First, uncertainty occurred more than other affective states in our prior ITSPOKE dialogues [14]. Second, uncertainty is of primary interest to tutoring researchers due to its theorized relationship to learning (e.g. [6,8,9]). In [8], VanLehn et al. view uncertainty and incorrectness as signalling "learning impasses": opportunities for the student to learn the material about which s/he is uncertain or incorrect. From this view we derived a specific uncertainty adaptation hypothesis to test in a controlled experiment: Responding to uncertainty *in the same way as* incorrectness will improve student performance, by providing students with the knowledge needed to resolve their uncertainty impasses.

Implementing this adaptation involved changing the next state transitions in the finite state dialogue manager; instead of transitioning based only on the correctness of the answer, the transition is based on the answer's combined correctness and uncertainty value. More specifically, our uncertainty adaptation consisted of treating all uncertain+correct answers as if they were incorrect (note that uncertain+incorrect answers are already treated as incorrect).

[1] [10] describes the resulting publicly available Uncertainty Corpus in detail.

[2] We use "affect" to cover emotions and attitudes. Some argue for separating them, but some speech researchers find the narrow sense of "emotion" too restrictive since it excludes speech where emotion is not full-blown, including arousal and attitude [13]. Some tutoring researchers also combine emotion and attitude (e.g. [5,7]).

For an initial investigation into the impact of this adaptation on student performance, we implemented it in a Wizard of Oz (WOZ) version of ITSPOKE that tutors only one physics problem (as opposed to five). In this WOZ, a few system components are replaced by a human "wizard": The wizard performs speech recognition, correctness annotation, and uncertainty annotation, for each student answer. In this way, we tested the adaptation hypothesis without any potentially negative impact of automated versions of these tasks. Upon hearing each student answer, the Wizard annotates if it is correct or uncertain. These distinctions are binary: a "correct" answer may be partially or fully correct, and a "nonuncertain" answer may be certain or neutral for certainty.[3]

2.2 Experimental Design

The experiment had 3 conditions, designed to test whether our uncertainty adaptation improved student performance. For use in these 3 conditions, the dialogue manager was parameterized, so that it could adapt contingently on the student state of uncertain+correct as discussed above, or randomly, or not at all.

In the **experimental condition**, the dialogue manager adapted to uncertainty by treating all uncertain+correct student answers as incorrect.

In the **normal control condition,** the dialogue manager did not adapt to uncertainty (it was merely logged); it treated only incorrect answers as incorrect. In other words, this condition corresponds to the original system.

In the **random control condition**, the dialogue manager did not respond to uncertainty (it was merely logged), but it did treat a percentage of random correct answers as incorrect. This condition was included to control for the additional tutoring dialogue given to students in the experimental condition. The percentage was toggled to be statistically the same as the percentage of answers adapted to in the experimental condition (approximately 20%).

Note that the uncertainty adaptation for uncertain+correct answers was always preceded by the correctness feedback "Okay"; all other answers received the regular ITSPOKE correctness feedback. "Okay" was intended to mirror this feedback behavior. In particular, ITSPOKE is fully automated and uses somewhat vague correctness feedback (e.g. "Fine" and "Well..." for recognized correct and incorrect answers, respectively) because it may misrecognize an answer.

Figures 1-2 illustrate how system responses differ across conditions. In both figures, $\mathbf{STUDENT}_1$ is *uncertain+correct*. In Figure 1 (experimental), $\mathbf{TUTOR}_2$ says "Okay", then asks an additional remediation question to remediate the student's uncertainty. In Figure 2 (normal control),$\mathbf{TUTOR}_2$ says "Good" and the remediation question is not asked. Section 3 discusses the tutor turn labels.

The experimental procedure was as follows. Each subject: 1) Read a small document of background physics material; 2) took a pretest of 7 fill in the blank

[3] In prior ITSPOKE corpora, this wizard displayed interannotator agreement of 0.85 Kappa on labeling binary correctness, and 0.62 Kappa on binary uncertainty [14]. Other possible approaches to labeling student affect include self-report (e.g. [4]).

TUTOR$_1$: What will the velocity of the object be a second after that (where the initial velocity is 9.8m/s and the acceleration is 9.8m/s^2)? *[Depth 1]*

STUDENT$_1$: Uh. 19.6 m/s? *[uncertain+correct]*

> **TUTOR**$_2$: Okay. As we have seen, if a falling object has an acceleration of 9.8m/s^2, that means that its velocity changes by 9.8m/s every second. So if a second after it began falling the object's velocity is 9.8m/s, a second later its velocity will be 9.8m/s+9.8m/s=19.6m/s. So what will its velocity be a second after it is 19.6m/s? *[Depth 2+]*
>
> **STUDENT**$_2$: Um. I'm thinking. 29.4 m/s. *[nonuncertain+correct]*

TUTOR$_3$: Right. So at every point in time during the fall of the man and his keys, how do their velocities compare with each other? *[Depth 1]*

Fig. 1. Corpus Excerpt: Experimental Condition

TUTOR$_1$: What will the velocity of the object be a second after that (where the initial velocity is 9.8m/s and the acceleration is 9.8m/s^2)? *[Depth 1]*

STUDENT$_1$: 19.6 m/s? *[uncertain+correct]*

TUTOR$_2$: Good. So at every point in time during the fall of the man and his keys, how do their velocities compare with each other? *[Depth 1]*

Fig. 2. Corresponding Corpus Excerpt: Normal Control Condition

questions; 3) used a web/voice interface to work through one physics *training* problem with a version of the WOZ (depending on condition); 4) took a posttest isomorphic to the pretest; 5) worked through a *test* problem isomorphic to the training problem with the *non-adaptive* WOZ (from the normal condition). Note that unlike the posttest, completing the test problem yielded a new dialogue.

Subjects were native English speakers who had not taken college physics. 60 subjects were randomly assigned to the 3 conditions (20 per condition), except conditions were gender-balanced. After the experiment, we found that 3 subjects in the experimental condition had no correct+uncertain answers and so never received the adaptation; 2 subjects in the random condition had no correct answers randomly selected for adaptation. These subjects were reclassified into the normal condition for our performance analysis.

3 Comparing Dialogue-Based Performance Metrics

We hypothesized that the training problem might be too short to yield significant differences between conditions in learning as measured by our pretest and posttest. This expectation was borne out; a two-way ANOVA with condition by repeated test measures design showed a significant main effect for test phase, ($F(1,57) = 33.919$, $p = 0.000$, $MSe = 0.032$), indicating students learned

overall, but there was no significant interaction effect between condition and test phase, indicating that amount of learning was not dependent on condition. One-way ANOVAs with post-hoc Tukey indicated no significant difference between conditions in raw (post-pre) or normalized ((post-pre)/(1-pre)) learning gain.

Thus, we used the test problem as an additional test of how the uncertainty adaptation in the training problem impacted student answers to the isomorphic questions in the test problem (where all students used the *non-adaptive* system, thereby receiving the same "test"). Below we analyze differences between conditions in dialogue-based performance metrics extracted from the test problem.

3.1 Comparing Impasse State Severities

In order to resolve a learning impasse, the student must first perceive that an impasse exists. Incorrectness and uncertainty differ in terms of this perception. Incorrectness simply indicates that the student has reached an impasse, while uncertainty - in a correct or incorrect answer - indicates that the student perceives s/he has reached an impasse. Based on this distinction, we associated each of our four answer combinations of uncertainty (**U**, **nonU**) and correctness (**I**, **C**) in the test problem with a scalar value from 3 to 0, as shown in Figure 3.

We hypothesized that these scalar values correspond to the severity of the student's current learning impasse state with respect to the test question, after receiving tutoring about the question in the training problem. Thus, 0 is a state in which the student is not experiencing an impasse, because s/he is correct and not uncertain about it. 3 is a state in which the student is experiencing the most severe type of impasse, because s/he is incorrect and not aware of it. 2 and 1 are states of lesser severity: the student is incorrect but aware that s/he might be, and the student is correct but uncertain about it, respectively.

Nominal State:	InonU	IU	CU	CnonU
Scalar State:	3	2	1	0
Severity Ranking:	most	less	least	none

Fig. 3. Different Impasse State Severities

After assigning a scalar state to each answer in the test problem, we computed a total and average impasse state severity per student. For example, suppose Figure 1 constituted our dataset for one student. The two student turns are labeled *uncertain+correct* and *nonuncertain+correct*, corresponding to scalar values 1 and 0, respectively. Thus the total = 1 (1+0), and the average = 0.5 (1/2).

We hypothesized that the experimental condition would show significantly lower total and average impasse severity in the test problem, because the uncertainty adaptation helped resolve more impasses during training. The "Means" columns in Table 1 show the means per condition. As expected, the experimental condition had lower total and average severity than the random condition, and random was lower than the normal condition. However, a one-way ANOVA with post-hoc Tukey showed no significant differences or trends ($p > 0.10$).

Table 1. Means and Correlations for Total and Average Impasse Severity

Metric	Means			Correlation (60)	
	Expmntl (17)	NormCtrl (25)	RandCtrl (18)	R	p
Tot. Impasse Severity	6.76	7.36	7.28	-0.38	0.003
Ave. Impasse Severity	0.38	0.42	0.41	-0.41	0.001

Despite this, we still hypothesized that lower impasse severities in the test problem are better, from a learning perspective. To support this, we computed a partial Pearson's correlation over all 60 students between both total and average impasse severity and posttest score, controlled for pretest score (pretest and posttest are significantly correlated in our data). The last two columns in Table 1 show the results. As shown, both total and average severity are significantly negatively correlated with learning, suggesting that lower impasse severities in the test problem are related to increased learning. We thus continue to use this hypothesis in our interpretation of results in the next sections.

3.2 Comparing Questions Originally Answered Correct+Uncertain

To further examine the impact of the uncertainty adaptation, we investigated student answers to those tutor questions that were asked in the training problem, answered as correct+uncertain, and then repeated in the test problem. In other words, we investigated student performance on the intended target of the uncertainty adaptation: the correct+uncertain (CU) answers. Note that these answers were all adapted to in the experimental condition, some were adapted to in the random condition, and none were adapted to in the normal condition.

The goal of our uncertainty adaptation was to increase correctness and decrease uncertainty in the test problem. In terms of these two dimensions combined, the goal was to decrease the frequency of the more severe nominal impasse states in Figure 3. Thus for each student's answers, we computed a total and percent of answers labeled with each (nominal) impasse severity (InonU, IU, CU, CnonU), as well as of correct (C) and nonuncertain (nonU) answers. For example, suppose both tutor questions in Figure 1 were originally answered CU in the training problem and are now repeated in the test problem. The totals then are: C=2, nonU=1, InonU=0, IU=0, CU=1, CnonU=1. The percents are: C=100%, nonU=50%, InonU=0%, IU=0%, CU=50%, CnonU=50%.

We hypothesized that the totals and percents in the experimental condition would be lower for InonU and IU, and higher for C, nonU, CU, and CnonU, because the uncertainty adaptation would have helped resolve impasses about these questions (or would have helped increase correctness and decrease uncertainty independently of each other). To test this hypothesis we ran a one-way ANOVA with post-hoc Tukey for each of the 12 metrics. Table 2 only shows metrics yielding significant differences or trends ($p<0.1$). The first column indicates these are answers to repeated questions originally answered CU (CU → ...).

Table 2. Means and Differences for Answers to Questions Originally Answered CU

Metric	Condition	Mean	Diff	p
Tot. CU → C	Expmntl	4.53	> NormCtrl	0.07
	NormCtrl	2.64		
	RandCtrl	5.11	> NormCtrl	0.01
Pct. CU → C	Expmntl	96.20%	> NormCtrl	0.09
	NormCtrl	76.50%		
	RandCtrl	91.06%		
Tot. CU → nonU	Expmntl	3.47		
	NormCtrl	2.32		
	RandCtrl	4.00	> NormCtrl	0.03
Tot. CU → CnonU	Expmntl	3.35		
	NormCtrl	2.20		
	RandCtrl	3.89	> NormCtrl	0.02

The remaining columns list the condition, its mean, the condition with which a difference is found, the direction of this difference (> or <), and its significance.

The first two results suggest that (significantly or as a trend) CU answers are more likely to stay correct in the test problem if they receive the uncertainty adaptation in the training problem. Put another way, CU answers are more likely to become incorrect during testing if the uncertainty adaptation is not received during training. The last two results suggest that the uncertainty adaptation reduces uncertainty in both the experimental and random conditions; however, only in the random condition do these results reach significance.

3.3 Comparing Answers at Different Dialogue Depths

We next tested whether the differences observed for answers to repeated questions generalized to all student answers in the test problem. However, one-way ANOVAs with post-hoc Tukey indicated no differences between conditions ($p > 0.10$) for any of the metrics (totals and percents for each nominal impasse state severity, for correct answers, and for uncertain answers).

We hypothesized that this lack of generalization might be due to the fact that student answers in remediation subdialogues can behave differently than those in the top-level dialogue, as we've shown in prior work [11]. As discussed in Section 2, the top-level dialogue is driven by correct answers to questions about the main problem topics, while a remediation subdialogue about a main topic is initiated by an incorrect answer to a top-level question. Thus as a final analysis, we distinguished these two answer types, which we refer to as "Depth 1" and "Depth 2+" answers. We computed the same metrics as above for each answer type and ran a one-way ANOVA with post-hoc Tukey for each metric. We found a trend for more Depth2+ answers to be CU in the experimental condition, as compared to the normal condition. More generally, the means for total and percent CU at Depth2+ were highest in the experimental condition, and lowest in the normal control condition. These results thus suggest that the uncertainty

adaptation helped increase correctness, but did not help decrease uncertainty, specifically regarding remediation questions. We hope to find firmer evidence of this when we repeat this type of analysis using data from the ongoing study discussed in Section 5.

4 Discussion and Related Work

Overall, our results in this paper are encouraging but inconclusive as to the benefit of our uncertainty adaptation on student performance. We hypothesize that two experimental design issues may have prevented larger differences between conditions. First, the training problem was likely too short. On average, it lasted 15 minutes, contained 20 student turns, and only 4 student turns on average received the adaptation in the experimental and random conditions. Second, the correctness feedback, "Okay", which preceded the uncertainty adaptation, was likely too vague. During the experiment, the wizard observed that uncertain+correct students were often confused by this feedback. We believe that the vagueness of "Okay" may have left these uncertain students ignorant as to whether their answer was correct. This vagueness may have been less noticeable to the random students, because roughly half of the time they were not uncertain when receiving the adaptation. This may explain why our analyses show little reduction in uncertainty in the experimental condition. Although resolving these issues should yield larger performance increases in the experimental condition, it still may not tease apart differences with the random condition. For one thing, *some* CU answers in the random condition receive the adaptation. A solution might be to only adapt to CnonU answers randomly; however, this too might benefit performance, by increasing the certainty of those answers (i.e., a CnonU answer may be neutral or certain). We assume it would not benefit performance to adapt to *every* correct answer, as this gives an identical response to incorrect and correct answers (except for correctness feedback).

Another complication is that it is not clear what is the best way to handle the fact that not all subjects in the two adaptive conditions actually received the adaptation. Although we moved into the normal condition the 5 subjects who didn't receive the adaptation, this is not necessarily the best solution because it can introduce sample bias; however, note that both before and after moving the subjects, the conditions had no significant difference in the total number or percent of correct answers in the test problem. Alternative approaches are also problematic. Removing the 5 subjects, as in [10], can also bias the samples. Retaining the subjects can yield ambiguous performance metrics. For example, for these 5 subjects, the metric %CU → CnonU would have to be set to 0 or left undefined because the denominator is 0 (# training CU), but if set to 0, then the value has another interpretation where this denominator is nonzero but the numerator is 0 (# training CU → testing CnonU). Note finally that if we use the Bonferroni correction, then the p-value required for a trend in Table 2 is 0.1/12 = 0.01. While this corrects for spurious results due to chance (type I errors), it can allow actual results to be overlooked (type II errors). We thus emphasize

that our results are exploratory and suggest specific hypotheses to be tested in our performance analysis of a larger experiment now underway (Section 5).

Determining when to adapt based on uncertainty is still an open question. To our knowledge only one other controlled experiment has tested uncertainty adaptations in spoken dialogue tutoring. In [5], Pon-Barry et al. implemented and evaluated two human tutor responses to uncertain answers (correct and incorrect) in the SCoT-DC tutor. In their "random" condition, the adaptations were used after all answers. They found significantly increased learning in this random condition as compared to a normal condition, but not in the experimental condition, where the adaptations were used only after uncertainty. Although most other work targeting uncertainty in the tutoring system community has involved correlational studies (e.g. [6]), there are other examples of adaptive tutoring systems developed or in development, which recognize affect and respond with various forms of empathy or politeness (e.g. [2,3,15]).

5 Conclusion and Current Directions

We presented one of the first experimental evaluations of student performance in a dialogue-based tutoring system that automatically adapts to student uncertainty. Our performance metrics include correctness, uncertainty, and learning impasse severity, which is a novel metric combining these two dimensions. These were measured in a test problem dialogue after the training dialogue. Though not conclusive, our results suggest that the uncertainty adaptation does have a positive benefit on student performance. In particular, correct+uncertain answers are more likely to become incorrect in the test problem if the uncertainty adaptation is not received during training, but only in the random condition are these answers also more likely to become nonuncertain. While learning impasse severity didn't differ significantly across conditions, it did significantly negatively correlate with student learning.

We hypothesized that two experimental design issues may have prevented more performance benefits of the uncertainty adaptation: short tutoring treatment and vague correctness feedback. We are now conducting a larger version of this experiment that resolves these issues. For this new experiment, we have implemented the uncertainty adaptation for all five ITSPOKE physics problems (rather than one); students are tutored for approximately an hour before taking the posttest, and thus are more likely to benefit from the uncertainty adaptation. In addition, we have replaced the vague "Okay" feedback with phrases that are clearly indicative of correctness (e.g. "That's correct").

Acknowledgments

This work was done as part of the Pittsburgh Science of Learning Center, funded by National Science Foundation (NSF) award #SBE-0354420. This work is also funded by NSF awards #0631930 & #0428472. We thank the ITSPOKE Group.

References

1. Workshop on Modeling and Scaffolding Affective Experiences to Impact Learning: Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education (AIED), Marina Del Ray, CA, Online proceedings (July 2007), http://www.informatics.sussex.ac.uk/users/gr20/aied07/index.html
2. Wang, N., Johnson, W., Rizzo, P., Shaw, E., Mayer, R.: Experimental evaluation of polite interaction tactics for pedagogical agents. In: Proceedings of Intelligent User Interface Conference (IUI), pp. 12–19 (2005)
3. Hall, L., Woods, S., Sobral, D., Paiva, A., Dautenhahn, K., Wolke, D., Newall, L.: Designing empathic agents: Adults vs. kids. In: Proceedings of the Intelligent Tutoring Systems Conference (ITS), Maceio, Brazil, pp. 604–613 (2004)
4. McQuiggan, S., Mott, B., Lester, J.: Modeling self-efficacy in intelligent tutoring systems: An inductive approach. User Modeling and User-Adapted Interaction (UMUAI) 18(1-2), 81–123 (2008)
5. Pon-Barry, H., Schultz, K., Bratt, E.O., Clark, B., Peters, S.: Responding to student uncertainty in spoken tutorial dialogue systems. International Journal of Artificial Intelligence in Education 16, 171–194 (2006)
6. Craig, S., Graesser, A., Sullins, J., Gholson, B.: Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. Journal of Educational Media 29(3), 241–250 (2004)
7. Bhatt, K., Evens, M., Argamon, S.: Hedged responses and expressions of affect in human/human and human/computer tutorial interactions. In: Proceedings of Cognitive Science (CogSci), Chicago, USA, pp. 114–119 (2004)
8. VanLehn, K., Siler, S., Murray, C.: Why do only some events cause learning during human tutoring? Cognition and Instruction 21(3), 209–249 (2003)
9. Kort, B., Reilly, R., Picard, R.: An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In: Okamoto, T., Hartley, R., Kinshuk, J., Klus, P. (eds.) Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges, Madison, WI, pp. 43–48 (2001)
10. Forbes-Riley, K., Litman, D., Silliman, S., Purandare, A.: Uncertainty corpus: Resource to study user affect in complex spoken dialogue systems. In: Proceedings 6th Language Resources and Evaluation Conference (LREC), Marrakech, Morocco (May 2008)
11. Forbes-Riley, K., Rotaru, M., Litman, D.: The relative impact of student affect on performance models in a spoken dialogue tutoring system. User Modeling and User-Adapted Interaction 18(1-2), 11–43 (2008)
12. VanLehn, K., Jordan, P.W., Rosé, C.P., Bhembe, D., Böttner, M., Gaydos, A., Makatchev, M., Pappuswamy, U., Ringenberg, M., Roque, A., Siler, S., Srivastava, R., Wilson, R.: The architecture of Why2-Atlas: A coach for qualitative physics essay writing. In: Proceedings of Intelligent Tutoring Systems (2002)
13. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Communication 40(1-2), 5–32 (2003)
14. Forbes-Riley, K., Litman, D.: Analyzing dependencies between student certainness states and tutor responses in a spoken dialogue corpus. In: Dybkjaer, L., Minker, W. (eds.) Recent Trends in Discourse and Dialogue, pp. 275–304. Springer (2008)
15. Burleson, W., Picard, R.: Affective agents: Sustaining motivation to learn through failure and a state of stuck. In: Social and Emotional Intelligence in Learning Environments Workshop at the Intelligent Tutoring Systems Conference (ITS), Maceio, Brazil (2004)