Exploring the Relationships between Design, Students' Affective States, and Disengaged Behaviors within an ITS

Lakshmi S. Doddannara[1], Sujith M. Gowda[1], Ryan S.J.d. Baker[2], Supreeth M. Gowda[1], and Adriana M.J.B. de Carvalho[3]

[1] Worcester Polytechnic Institute, Worcester MA 01609
[2] Teacher's College, Columbia University, New York NY 10027
[3] Carnegie Mellon University, Pittsburgh, PA 15213
{Lakshmi1023,dikajoazeirodebaker}@gmail.com,
{sujithmg,smgowda}@wpi.edu,
baker2@exchange.tc.columbia.edu

Abstract. Recent research has shown that differences in software design and content are associated with differences in how much students game the system and go off-task. In particular the design features of a tutor have found to predict substantial amounts of variance in gaming and off-task behavior. However, it is not yet understood how this influence takes place. In this paper we investigate the relationship between a student's affective state, their tendency to engage in disengaged behavior, and the design aspects of the learning environments, towards understanding the role that affect plays in this process. To investigate this question, we integrate an existing taxonomy of the features of tutor lessons [3] with automated detectors of affect [8]. We find that confusion and frustration are significantly associated with lesson features which were found to be associated with disengaged behavior in past research. At the same time, we find that the affective state of engaged concentration is significantly associated with features associated with lower frequencies of disengaged behavior. This analysis suggests that simple re-designs of tutors along these lines may lead to both better affect and less disengaged behavior.

Keywords: Educational Data Mining, Intelligent Tutoring System, design features, affect, Gaming the System, Off-task behavior.

1 Introduction

There has been considerable research into students' disengaged behaviors in intelligent tutoring systems over the last few years [6, 7, 10, 11, 13, 15, 21, 29, 32]. This work has generally found that a range of disengaged behaviors are associated with negative learning outcomes, including both gaming the system and off-task behavior [cf. 1, 15, 30].

Early work on why students became disengaged investigated whether fairly non-malleable factors such as goal orientation or motivation could predict disengaged behaviors [e.g. 10, 11]. However, recent research has suggested that differences in the

K. Yacef et al. (Eds.): AIED 2013, LNAI 7926, pp. 31–40, 2013.

design of intelligent tutoring systems can also have substantial impacts on student engagement. Relatively simple aspects of design such as the concreteness of problem scenarios and hints were found to predict a considerable proportion of the variance in gaming the system among a group of students using Cognitive Tutor Algebra over the course of a year [6]. Off-task behavior has also been found to vary according to design features such as presence or absence of problem scenarios [3]. These findings suggest that design aspects of tutor lessons may play a significant role in influencing the prevalence of disengaged behavior.

However, we do not yet understand the mechanisms through which differences in the design of tutor lessons may influence disengaged behavior. One mechanism hypothesized in those earlier papers was that affect might be mediating the relationship between tutor design and disengaged behavior. There is evidence for reasonably strong relationships between affect and disengaged behavior. Research in Aplusix and The Incredible Machine (an ITS and a puzzle game) found that boredom preceded and co-occurred with a student's choice to game the system [7]. Boredom has also been found to precede off-task behavior [9] and off-task behavior within the learning environment (also called WTF/"without thinking fastidiously" behavior) within intelligent tutoring systems [32]. There is also evidence that boredom leads to future off-task behavior, within both the Chemistry Virtual Laboratory [9] and Science ASSISTments [22]. However, it is not yet known how strong the relationships are between intelligent tutor design features and affect.

Understanding the factors leading to differences in affect is important by itself as well. There is increasing evidence that differences in affect during use of educational software can have a substantial impact on learning. Craig and colleagues [16] investigated the relationships between learning gains and affect state and found that confusion and flow were positively associated with learning gains but boredom was negatively associated with learning. Pardos and colleagues [30] also found that affect in intelligent tutors can predict not just local learning, but longer-term learning outcomes (state standardized exam scores) as well, specifically finding that boredom is negatively associated with longer-term learning outcomes while engaged concentration (e.g. flow) and frustration were positively associated with learning gains. Evidence in that paper suggested that the context of affect matters more than the overall prevalence, with the relationship between boredom and learning outcomes reversing and becoming positive if the boredom occurs during scaffolding. Other work has suggested that the duration of affect also matters, with brief confusion correlating positively with learning but lengthy confusion correlating negatively with learning [26]. Flow/engaged concentration has also been shown to be associated with longer-term engagement with specific domains [17] One possible explanation for this finding is that positive affect may lead to increased situational interest [23], which in turn has been theorized to lead to greater long term personal interest in the content domain [25].

Given the relationship between disengaged behavior and affect, and the importance of affect in general, it may be worth considering the ways in which tutor design features drive not just disengaged behaviors, but affect as well. In this paper we study the

relationships between these three factors. We use an existing taxonomy of the features of tutor lessons [6] to express the differences between lessons. Taxonomies of this nature, also referred to as "design pattern languages" [34], can be useful tools for studying and understanding design. We integrate data from the application of this taxonomy to a set of lessons from an algebra tutor, with predictions from previously published automated detectors of affect [8] and disengaged behaviors [4, 5]. We then conduct correlation mining (with post-hoc controls) to study the relationships between these variables.

2 Data Set

Data was obtained from the PSLC DataShop (dataset: Algebra I 2005-2006 Hampton Only; this data set was chosen because it is readily available in the DataShop and has been studied in other research as well), for 58 students' use of Cognitive Tutor Algebra during an entire school year. A full description of the Cognitive Tutor used in this study can be found in [24]. The data set was composed of approximately 437,000 student transactions (entering an answer or requesting help) in the tutor software. All of the students were enrolled in algebra classes in one high school in the Pittsburgh suburbs which used Cognitive Tutors two days a week, as part of their regular mathematics curriculum. None of the classes were composed predominantly of gifted or special needs students. The students were in the 9th and 10th grades (approximately 14-16 years old). The Cognitive Tutor Algebra curriculum involves 32 lessons, covering a complete selection of topics in algebra, including formulating expressions for word problems, equation solving, and algebraic function graphing.

Data from 10 lessons was eliminated from consideration, to match the original analysis of this data in [6], where the relationship between tutor design and gaming the system was studied. In that original study, lessons were eliminated due to having insufficient data to be able to conduct a sufficient number of text replays to effectively measure gaming the system. On average, each student completed 9.9 tutor lessons (among the set of lessons considered), for a total of 577 student/lesson pairs.

3 Method

In describing the methods sections, first we will describe taxonomic feature generation process and then describe affect detection process used to build machine learned affect models which were in-turn used in this analysis to obtain affect predictions.

3.1 The Cognitive Tutor Lesson Variation Space (CTLVS)

The enumeration of the ways that Cognitive Tutor lessons can differ from one another was originally developed in [6]. This enumeration, in its current form, is called the Cognitive Tutor Lesson Variation Space version 1.2 (CTLVS1.2). The CTLVS was

developed by a six member design team with diverse expertise, including three Cognitive Tutor designers (with expertise in cognitive psychology and artificial intelligence), a researcher specializing in the study of gaming the system, a mathematics teacher with several years of experience using Cognitive Tutors in class, and a designer of non-computerized curricula who had not previously used a Cognitive Tutor.

During the first step of the design process, the six member design team generated a list with 569 features. In the next step a list of criteria for features that would be worth coding, were developed. Finally the list was narrowed down to a more tractable size of 79 features. Inter-rater reliability checks were not conducted, owing to the hypothesis-generating nature of this study. Then CTLVS1 was labeled with reference to the 21 lessons studied in this paper by a combination of educational data mining and hand coding by the educational designer and mathematics teacher. The 10 features among 79 within the CTLVS1.1 which were significant predictors of disengaged behaviors in [3, 6] are shown in Table 1.

After initial publication of the results [e.g. 3, 6], using the CTLVS 1.1, additional coding was conducted by the gaming the system researcher and the designer of non-computerized curricula resulting in the addition of 5 more features, shown in Table 2. This produced a total of 84 quantitative and binary features within the CTLVS1.2.

Table 1. Design features which were significant predictors of disengaged behaviors in [3, 6]

1. Lesson is an equation-solver lesson, where a student is given an equation to solve mathematically (with no story problem)
2. Avg. amount that reading on-demand hints improves performance on future opportunities to use skill (using model from [12])
3. % of hint sequences with final "bottom-out" hint that explicitly tells student what to enter [cf. 1]
4. Reference in problem statement to interface component that does not exist (ever occurs)
5. Not immediately apparent what icons in toolbar mean
6. Hint requests that student perform some action
7. % of hints that explicitly refer to abstract principles
8. % of problem statements that use same numeric value for two constructs
9. % of problem statements with text not directly related to problem-solving task (typically included to increase interest)
10. Any hint gives directional feedback (example: "try a larger number")

3.2 Affect Detection Process

In order to study the relationship between students' affect and tutor design, we used previously developed detectors of student affect within Cognitive Tutor Algebra [cf. 8]. See [8] for a full discussion of the detectors. Unlike many of the pioneering efforts to detect student affect in intelligent tutoring systems [2, 18, 27], this work does not

make use of any visual, audio or physiological sensors such as webcams, pressure sensing keyboard and mice, pressure sensitive seat pads and back pads, or wireless conductance bracelets in detecting affect. Instead, affect is detected solely from log files, supporting scalable analyses. These affect detectors were originally developed by labeling a set of students' affective states with field observations and then using those labels to create machine-learned models which automatically detect the student's affective state. Affect detectors were developed for the states of boredom, confusion, frustration, and engaged concentration (the affect associated with the flow state [cf. 7]. A separate detector was developed for each affective state. The goodness of the detectors (under student-level cross-validation) is given in Table 3; the detectors agree with human coders approximately half as well as human coders agree with each other. Note that the A' values for the models are lower than presented in the original paper [8]. This is because the implementation of AUC in RapidMiner 4.6 [28] was used to compute the A' values. This implementation has a bug, where estimates of A' are inflated, if multiple data points have the same confidence. In this paper we report estimates computed through directly computing the A'/Wilcoxon statistic, which is more computationally intensive but mathematically simpler (involving a set of pairwise comparisons rather than integrating under a complex function), using the code at http://www.columbia.edu/~rsb2162/edmtools.html .

Table 2. The design features added in CTLVS1.2

1. % of hints with requests for students with politeness indicators
2. % of scenarios with text not directly related to problem-solving task
3. Maximum number of times any skill is used in problem
4. Average number of times any skill is used in problem
5. Were any of the problem scenarios lengthy and with extraneous text? (Long Extraneous Text)

Table 3. Goodness of the affect models [cf. 8]

Affect	**Algorithm**	**Kappa**	**A'**
Engaged Concentration	K*	0.31	0.67
Boredom	Naïve Bayes	0.28	0.69
Confusion	JRip	0.40	0.71
Frustration	REPTree	0.23	0.64

To apply the machine-learned models to the data set used in this paper, we computed the features which were used in the models. The data was divided into "clips", of 20 second intervals of student behavior (the same grain-size used in the original observations which were used to build the detector), using the absolute time of each

student action. Next, the 15 features used in the detectors [cf. 8] were computed for each clip. Finally RapidMiner 4.6 [28] was used to load each of the affect models and then each of the affect models were applied on the algebra data set to obtain assessments of affect for each clip, which were then aggregated to compute each student's proportion of each affective state in each lesson.

4 Results

For each lesson in the data set, we computed values for each of the 84 taxonomical features discussed in the data section. The value of each taxonomic feature was then correlated to the proportion of each of the four affective states (engaged concentration, boredom, confusion and frustration) detected within the log data for the lesson. As this represents a substantial number of statistical analyses (84*4 = 336), we controlled for multiple comparisons. In specific, the analyses in this study utilize the false discovery rate (FDR) [14] paradigm for post-hoc hypothesis testing, using Storey's method [33]. This method produces a substitute or p-values, termed q-values, driven by controlling the proportion of false positives obtained via a set of tests. Whereas a p-value expresses that 5% of all tests may include false positives, a q-value indicates that 5% of significant tests may include false positives. As such, the FDR method does not guarantee each test's significance, but guarantees a low overall proportion of false positives. This avoids the substantial Type II errors (over-conservatism) associated with the better-known Bonferroni correction [see 31 for a discussion of current statistical thought on the Bonferroni correction]. The FDR calculations in the results section were made using the QVALUE software package [33] within the R statistical software environment.

Across the features, only the five following tutor design features achieved statistically significant correlation to any of the affective states.

1. Lesson is an Equation Solver lesson (Equation Solver)
2. % of problem statements with text not directly related to problem-solving task (Extraneous Text),
3. % of problem statements which involve concrete people/places/things (Concrete Problem Statements),
4. Were any of the problem scenarios lengthy and with extraneous text? (Long Extraneous Text)
5. Average percent error in problem (Percent Error)

Table 4 summarizes the results. Equation Solver was statistically significantly positively associated with Concentration, $r=0.728$, $t(1,19)=4.622$, $q<0.01$; on the other hand 2 of the features Concrete Problem Statements and Long Extraneous Text were statistically significantly negatively associated with Concentration; Concrete Problem Statements $r= -0.604$, $t(1,19)= -3.31$, $q=0.013$; Long Extraneous Text $r= -0.538$, $t(1,19)= -2.78$, $q=0.032$.

Table 4. Statistical Significant results with q-values from FDR analysis

Design Features	Affect	r	Q
Equation Solver	Concentration	0.728	<0.01
Extraneous Text	Confusion	0.787	<0.001
Concrete Problem Statements	Concentration	-0.604	0.013
Concrete Problem Statements	Confusion	0.644	<0.01
Long Extraneous Text	Concentration	-0.538	0.032
Long Extraneous Text	Confusion	0.716	<0.01
Percent Error	Frustration	-0.718	<0.01

Three of the features were statistically significantly positively associated with Confusion, Concrete Problem Statement $r=0.644$, $t(1,19)=3.67$, $q<0.01$; Long Extraneous Text $r=0.716$, $t(1,19)=4.47$, $q<0.0$; Extraneous Text $r=0.787$, $t(1,19)=5.56$, $q<0.001$.

Only one of the features, Percent Error was statistically significantly negatively associated with Frustration, $r= -0.718$, $t(1,19)= -4.5$, $q<0.01$.

None of the features showed significant association with Boredom. The strongest correlation was achieved by "Hint gives directional feedback (example: "try a larger number")", $r=0.50$, $t(1,19) = 2.5$, $q=0.30$. It is worth noting that the original p value, before post-hoc correction, was $p=0.02$; hence, it may be worth considering this feature in future research, but there is insufficient evidence to make a conclusive inference about it at this point.

In terms of past features associated with gaming (in [6], it was hypothesized that this relationship was mediated by boredom), boredom appeared to be weakly correlated with Extraneous Text $r=0.160$, $t(1,19) = 0.71$, $q=0.78$ and Long Extraneous Text $r=0.264$, $t(1,19)=1.19$, $q=0.64$ and appeared to be moderately correlated with Concrete Problem Statements, $r=0.335$, $t(1,19)= 1.55$, $q=0.64$. None of these relationships, however, would be statistically significant even without post-hoc controls.

5 Discussion and Conclusions

The result here suggests that there are significant relationships between affect state of students, and the taxonomic features of an intelligent tutoring system. Five out of 84 taxonomic features were found to be statistically significantly associated with three affective states, engaged concentration, frustration, and confusion. These findings correspond in interesting ways to prior results regarding the relationship between disengaged behaviors and these same taxonomic features [cf. 3, 6].

Students were found to be concentrating significantly more during equation-solver lessons. These same lessons have also been found to be associated with a lower degree of off-task behavior and gaming the system in the previous research [3, 6].

We also found that students' concentration was reduced when the student encountered lessons with substantial extraneous text, as well as or problem statements and scenarios

with concrete people, places or things. These same features were also associated with increased confusion. These are somewhat surprising findings, as extraneous text was also associated with gaming the system in earlier research [6]. Since gaming is thought to be negatively associated with engaged concentration [7], it is surprising that the same features of an interface are associated both with gaming and less engaged concentration. This finding clearly calls for greater research to understand its full implications.

At the same time, the connection between substantial extraneous text and concrete scenarios, and confusion, accords well to past findings in other contexts. The details in these long concrete scenarios could be considered "seductive details" – details which draw student attention away from the content. Seductive details have been found to be associated with poorer learning in laboratory studies [20]; the initial interpretation of [6] seemed to contradict this finding, but our results here seem more in keeping with it. Of course, it also may be that tutor designers have chosen (whether consciously or not) to increase the complexity of the scenarios when material is more confusing; as such, it would take an experimental study to be fully confident of the hypothesis generated here.

One unexpected finding was negative correlation between percent error and frustration, which should be investigated further. In a different intelligent tutor, frustration was found to be positively correlated with learning, suggesting that frustration's role in learning may be somewhat different than typically hypothesized [cf. 30].

Another surprising finding is that none of the taxonomic features were significantly associated with boredom, a persistent affect state within several types of learning environments [7]. We had earlier hypothesized that the negative relationship between gaming and lengthier scenarios would be mediated by boredom [e.g. 6], a finding not obtained here. Though we found some appearance of correlation between boredom and lengthier scenarios as well as other features known to be associated with gaming, these associations were not significant even without taking post-hoc adjustment into account, suggesting that it is unlikely that boredom is a key mediator between these tutor design features and gaming the system.

One valuable area of future work would be to extend the research here to additional affective states, such as delight, disgust, and anxiety. The affective states chosen in this research were selected because relevant detectors already existed, and because these states have high theoretical importance and/or are known to correlate with differences in learning outcomes and engagement; extending to additional affective states would help to create a fuller picture of the relationships between affect and tutor design.

One of the final things that can be noted from this analysis is that the designs of educational interfaces can have a considerable impact on student affect. Although only a relatively small number of relationship remained significant after post-hoc testing, it is worth noting that the conservatism of post-hoc approaches meant that the relationships that remained significant had extremely high correlations (in the 0.7 range). This finding implies that relatively small differences in intelligent tutors may result in substantial impacts on student experiences.

Acknowledgments. The authors thank the Pittsburgh Science of Learning Center (National Science Foundation) via grant "Toward a Decade of PSLC Research", award number SBE- 0836012.

References

1. Aleven, V., McLaren, B., Roll, I., Koedinger, K.R.: Toward tutoring help seeking: Applying cognitive modeling to meta-cognitive skills. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) ITS 2004. LNCS, vol. 3220, pp. 227–239. Springer, Heidelberg (2004)
2. Arroyo, I., Woolf, B.P., Cooper, D., Burleson, W., Muldner, K., Christopherson, R.: Emotion Sensors Go To School. In: Proc. of the 14th International Conference on Artificial Intelligence in Education (2009)
3. Baker, R.S.J.d.: Differences Between Intelligent Tutor Lessons, and the Choice to Go Off-Task. In: Proc. of the 2nd Int'l. Conference on Educational Data Mining, pp. 11–20 (2009)
4. Baker, R.S.J.d.: Modeling and understanding students' off-task behavior in intelligent tutoring systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1059–1068 (2007)
5. Baker, R.S.J.d., de Carvalho, A.M.J.A.: Labeling Student Behavior Faster and More Precisely with Text Replays. In: Proceedings of the 1st International Conference on Educational Data Mining, pp. 38–47 (2008)
6. Baker, R.S.J.d., de Carvalho, A.M.J.A., Raspat, J., Aleven, V., Corbett, A.T., Koedinger, K.R.: Educational Software Features that Encourage and Discourage "Gaming the System". In: Proc. of the 14th Int'l. Conf. on Artificial Intelligence in Education, pp. 475–482 (2009)
7. Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C.: Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. International Journal of Human-Computer Studies 68(4), 223–241 (2010)
8. Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., Rossi, L.: Sensor-free automated detection of affect in a Cognitive Tutor for Algebra. In: Proceedings of the 5th International Conference on Educational Data Mining, pp. 126–133 (2012)
9. Baker, R.S.J.d., Moore, G.R., Wagner, A.Z., Kalka, J., Salvi, A., Karabinos, M., Ashe, C.A., Yaron, D.: The Dynamics Between Student Affect and Behavior Occuring Outside of Educational Software. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part I. LNCS, vol. 6974, pp. 14–24. Springer, Heidelberg (2011)
10. Baker, R.S.J.d., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., Koedinger, K.: Why Students Engage in "Gaming the System" Behavior in Interactive Learning Environments. Journal of Interactive Learning Research 19(2), 185–224 (2008)
11. Beal, C.R., Qu, L., Lee, H.: Mathematics motivation and achievement as predictors of high school students' guessing and help-seeking with instructional software. Journal of Computer Assisted Learning 24(6), 507–514 (2008)
12. Beck, J.E., Chang, K.-M., Mostow, J., Corbett, A.: Does Help Help? Introducing the Bayesian Evaluation and Assessment Methodology. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 383–394. Springer, Heidelberg (2008)
13. Beck, J.: Engagement tracing: using response times to model student disengagement. In: Proc. of 12th Int'l Conference on Artificial Intelligence in Education, pp. 88–95 (2005)
14. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57, 289–300 (1995)
15. Cocea, M., Hershkovitz, A., Baker, R.S.J.d.: The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate? In: Proceedings of the 14th International Conference on Artificial Intelligence in Education, pp. 507–514 (2009)
16. Craig, S.D., Graesser, A.C., Sullins, J., Gholson, B.: Affect and learning: an exploratory look into the role of affect in learning. J. of Educational Media 29, 241–250 (2004)
17. Csikszentmihalyi, M., Schneider, B.: Becoming Adult. Basic Books, New York (2001)

18. D'Mello, S.K., Graesser, A.C.: Multimodal semiautomated affect detection from conversational cues, gross body language, and facial features. User Modeling and User-adapted Interaction 20(2), 147–187 (2010)
19. D'Mello, S.K., Taylor, R., Graesser, A.C.: Monitoring Affective Trajectories during Complex Learning. In: Proc. of the 29th Annual Conf. of the Cognitive Science Society, pp. 203–208 (2007)
20. Harp, S.F., Mayer, R.E.: How seductive details do their damage: a theory of cognitive interest in science learning. Journal of Educational Psychology 90, 414–434 (1998)
21. Hastings, P., Arnott-Hill, E., Allbritton, D.: Squeezing out gaming behavior in a dialog-based ITS. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010, Part I. LNCS, vol. 6094, pp. 204–213. Springer, Heidelberg (2010)
22. Hershkovitz, A., Baker, R.S.J.d., Gobert, J., Nakama, A.: A Data-driven Path Model of Student Attributes, Affect, and Engagement in a Computer-based Science Inquiry Microworld. In: Proceedings of the International Conference on the Learning Sciences (2012)
23. Hidi, S., Anderson, V.: Situational interest and its impact on reading and expository writing. In: Renninger, K.A., Hidi, S., Krapp, A. (eds.) The Role of Interest in Learning and Development, pp. 215–238. Erlbaum, Hillsdale (1992)
24. Koedinger, K., Corbett, A.: Cognitive Tutors: Technology bringing learning science to the classroom. In: Sawyer, K. (ed.) The Cambridge Handbook of the Learning Sciences, pp. 61–78. Cambridge University Press (2006)
25. Krapp, A.: Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. Learning and Instruction 12(4), 383–409 (2002)
26. Lee, D.M.C., Rodrigo, M.M.T., Baker, R.S.J.d., Sugay, J.O., Coronel, A.: Exploring the Relationship between Novice Programmer Confusion and Achievement. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part I. LNCS, vol. 6974, pp. 175–184. Springer, Heidelberg (2011)
27. Litman, D.J., Forbes-Riley, K.: Recognizing Student Emotions and Attitudes on the Basis of Utterances in Spoken Tutoring Dialogues with both Human and Computer Tutors. Speech Communication 48(5), 559–590 (2006)
28. Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T.: Yale: rapid prototyping for complex data mining tasks. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 935–940 (2006)
29. Murray, R.C., VanLehn, K.: Effects of dissuading unnecessary help requests while providing proactive help. In: Proc. of the Int'l Conf. on Artificial Intelligence in Education (2005)
30. Pardos, Z., Baker, R.S.J.d., San Pedro, M.O.Z., Gowda, S.M., Gowda, S.: Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In: To Appear in Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (in press)
31. Perneger, T.V.: What's wrong with Bonferroni adjustments. British Medical Journal 316, 1236–1238 (1998)
32. Sabourin, J., Rowe, J.P., Mott, B.W., Lester, J.C.: When Off-Task is On-Task: The Affective Role of Off-Task Behavior in Narrative-Centered Learning Environments. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS, vol. 6738, pp. 534–536. Springer, Heidelberg (2011)
33. Storey, J.D., Taylor, J.E., Siegmund, D.: Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. Journal of the Royal Statistical Society, Series B 66(1), 187–205 (2004)
34. Van Duyne, D.K., Landay, J.A., Hong, J.I.: The design of sites patterns for creating winning web sites, Upper Saddle River, NY (2008)