Towards an Understanding of Affect and Knowledge from Student Interaction with an Intelligent Tutoring System

Maria Ofelia Z. San Pedro[1], Ryan S.J.d. Baker[1], Sujith M. Gowda[2],Neil T. Heffernan[2]

[1]Teachers College, Columbia University, New York, NY
[2]Worcester Polytechnic Institute, Worcester, MA
mzs2106@tc.columbia.edu, baker2@exchange.tc.columbia.edu, mgsujith@gmail.com, nth@wpi.edu

ABSTRACT. Csikszentmihalyi's Flow theory states that a balance between challenge and skill leads to high engagement, overwhelming challenge leads to anxiety or frustration, and insufficient challenge leads to boredom. In this paper, we test this theory within the context of student interaction with an intelligent tutoring system. Automated detectors of student affect and knowledge were developed, validated, and applied to a large data set. The results did not match Flow theory: boredom was more common for poorly-known material, and frustration was common both for very difficult material and very easy material. These results suggest that design for optimal engagement within online learning may require further study of the factors leading students to become bored on difficult material, and frustrated on very well-known material.

Keywords: Affect Modeling, Prior Knowledge, Intelligent Tutoring System, Boredom, Frustration, Engaged Concentration.

1 Introduction

In recent years, substantial work has gone into increasing the sensitivity and responsiveness of intelligent tutoring systems (ITSs) to differences in student affect [10, 11]. One theory that has inspired design in education [cf. 28] is Csikszentmihalyi's Flow theory [8]. This theory details the attributes of optimal experience during activity, making a number of specific claims that can be investigated, tested, and leveraged within design when a person is engaged in an activity with clear goals, with immediate feedback, and when balance is achieved between the person's perception of task difficulty and perception of one's own skills to do the task [8]. Empirical work in classrooms using traditional approaches (e.g., not ITS) has found that high school students experience the highest engagement when students perceive both challenge and their skill as high [28]. Csikszentmihalyi [8, 9] also hypothesized that specific affective states (emotion in context [cf. 7]) emerge depending on the degree of challenge and skill that is present for an activity. His theory indicates that when an activi-

adfa, p. 1, 2013.
© Springer-Verlag Berlin Heidelberg 2013

ty is perceived to be too easy one becomes bored, and when the task is too difficult one gets anxious [8]. An additional hypothesis is that the same conditions that lead to anxiety also lead to frustration [13], implying that challenge is higher than skill, leading some researchers to use frustration rather than anxiety in applying Csikszentmihalyi's theory [cf. 20, 25].

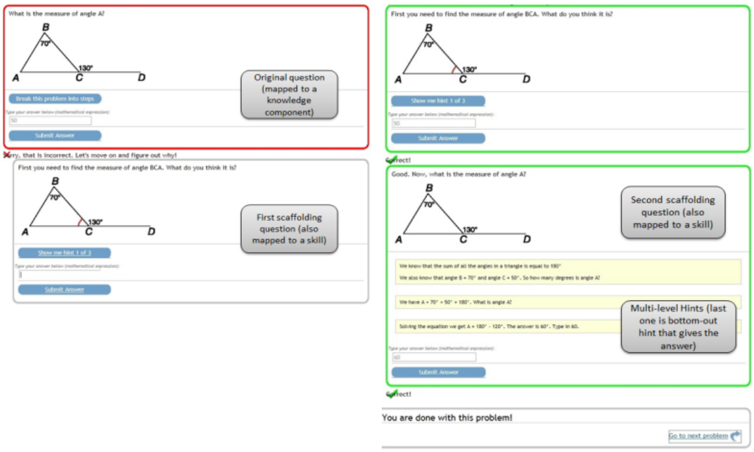
Flow theory, when applied to the context of education, asserts that a learning activity should be perceived as challenging but not too difficult [27]. As such, non-adaptive learning materials are likely to fail in producing flow for most students, as materials at a specific difficulty level are likely to be boring for students with higher skill, and frustrating for students with lower skill [cf. 26]. However, a learning system that accurately infers student skill – as modern intelligent tutoring systems do – may be able to specifically select problems of appropriate difficulty, in an attempt to balance challenge with skill level [18].

However, there is still not sufficient empirical evidence that Flow theory's account of the consequences of failing to achieve a balance between difficulty and skill are as predicted. In particular, recent research has suggested that boredom is often characteristic of the least successful students rather than students who have already achieved mastery [1, 7, 19]. This same research finds that frustration does not appear to be strongly connected with the poorest students [7, 22, 23]. These studies have the limitation of investigating these issues at a fairly coarse grain-size, looking solely at overall prevalence of affective states and long-term measures of learning. By studying these issues at a finer grain-size, we can understand these relationships better.

In this paper, we operationalize boredom, frustration, and engaged concentration during online learning in the fashion proposed in [3, 7]. In this paradigm, affective states are conceptualized as atomic and distinct from one another. Of particular importance to Flow theory are boredom [8, 15], frustration [13], and engaged concentration [cf. 3], which is the affect associated with Csikszentmihalyi's construct of flow but does not involve the inherent task-related aspects of flow – clear goals, immediate feedback, and balance between challenge and skill.

We conduct this research in a data set of 8,454 students learning online for a year apiece in the ASSISTment system [21], a free web-based tutoring system for middle school mathematics. Within ASSISTments, students complete mathematics problems and are formatively assessed – providing detailed information on their knowledge to their teachers – while being assisted with scaffolding, help, and feedback. Items in ASSISTments are designed to correspond to the skills and concepts taught in relevant state standardized examinations. Teachers have the ability to assign students questions on a particular skill and typically select the problems or problem sets their students receive (though mastery learning can also be activated by the teacher for some problem sets). As shown in Figure 1, the ASSISTment system provides feedback on incorrect answers. When a student answers a problem incorrectly, they are provided with scaffolding questions breaking the problem into its component steps. Hints are provided at each step and the student can ask for a bottom-out hint that eventually tells the answer.

Within this paper, we use automated detectors of student affect within the ASSISTment system (published in previous work [16]) to operationalize student affect within the ASSISTment system. These detectors, developed and validated using data from 229 students, are then applied to the full data set of 8,454 students. We combine these detectors with data from models of student knowledge in order to analyze the conditions under which each affective state occurs, and whether the relationship between affect and the difficulty of a problem for a specific student accords with Flow theory. We conclude with a discussion of potential implications for the design of interactive educational systems.



a) b)

Fig. 1. Example of an ASSISTment. a) If a student gets it incorrect, hints and scaffolding problems are there to aid the student in eventually getting the correct answer. b) Example of Scaffolding and Hints in an ASSISTment.

2 Measures Used

2.1 Affect Detectors

Within this paper, we leverage existing detectors of student affect within the ASSISTment system [16], to help us understand student affect across contexts. Detectors of three affective states are utilized: engaged concentration, boredom, and frustration. The detectors of engaged concentration and boredom used in this paper are identical to the detectors used in [16]. After publishing [16], we discovered a minor computation error in one of the features used in the frustration detector. Hence, a re-computed model is used here (the goodness of the detector is almost exactly identical between the [16] and this paper). Though anxiety plays a prominent role in Csikszentmihalyi's Theory of Flow, no detector of anxiety in ASSISTments was available, in part because anxiety has been observed so rarely in classroom use of intelligent tutoring systems as to not merit its own coding category [12, 14, 23].

These detectors were developed using a two-stage process: first, student affect was labeled for a sample of 3,075 field observations [cf. 3] of 229 students conducted by two coders using an Android app, and then those labels were used to create automated detectors that can be applied to log files at scale. An inter-rater reliability session was conducted, where the two coders coded the same student at the same time (they observed multiple students, but observed each student together). They conducted 51 simultaneous observations, achieving a Cohen's Kappa of 0.72, indicating agreement 72% better than chance. The detectors were created by synchronizing log files generated by the ASSISTments system with field observations conducted at the same time. To enhance scalability, only log data was used as the basis of the detectors, instead of using physical sensors (and indeed, the research presented in this paper could not have been conducted if physical sensors were used). The detectors were constructed using only log data from student actions within the software occurring at the same time as or before the observations. By using information only from before and during the observation, our detectors can be used for automated interventions, as well as the discovery with models analyses presented in this paper.

All of the affect detectors performed better than chance. Detector goodness within ASSISTments was at the high end of previous reports of published models inferring student affect in an ITS solely from log files [cf. 4, 5, 11, 24]. The best detector of engaged concentration involved the K* algorithm, achieving an A' of 0.678 and a Kappa of 0.358. The best boredom detector was found using the JRip algorithm, achieving an A' of 0.632 and a Kappa of 0.229. The best frustration detector achieved an A' of 0.681 and a Kappa of 0.301, using the J48 algorithm. These levels of detector goodness indicate models that are clearly informative, though there is still considerable room for improvement.

Within the original observations, boredom was observed 17.7% of the time, frustration was observed 4.4% of the time, and engaged concentration 53.0% of the time, with other affective states representing the remainder of student time. The detectors emerging from the data mining process had some systematic error in prediction, where the average confidence of the resultant models was systematically higher or lower than the proportion of the affective states in the original data set. This type of bias does not affect correlation to other variables since relative order of predictions is unaffected, but it can reduce model interpretability. To increase model interpretability, model confidences were rescaled to have the same mean as the original distribution, using linear interpolation. Rescaling the confidences this way does not impact model A' or Kappa, as it does not change the relative ordering of model assessments.

2.2 Prior Knowledge Assessment

Estimates of student knowledge were used as a proxy for Flow theory's "balance between challenge and skill." These estimates were computed using Bayesian Knowledge Tracing (BKT) [6], a model used in several ITSs to estimate a student's latent knowledge based on his/her observable performance. This model can predict how difficult the current problem will be for the current student, based on the skills required for that problem. As such, this model can implicitly capture the tradeoff between difficulty and skill for the current context. This model can inform us whether student skill is higher than current difficulty (resulting in a high probability of correctness), when current difficulty is higher than student skill (resulting in a low probability of correctness), and when difficulty and skill are in balance (medium probabilities of correctness). To assess student skill, BKT infers student knowledge by continually updating the estimated probability a student knows a skill every time the student gives a first response to a new problem. It uses four parameters, each estimated separately per skill: L_O, the initial probability the student knows the skill; T, the probability of learning the skill at each opportunity to use that a skill; G, the probability that the student will give a correct answer despite not knowing the skill; and S, the probability that the student will give an incorrect answer despite knowing the skill. In this model, the four parameters for each skill are held constant across contexts and students (variants of BKT relax these assumptions). BKT uses Bayesian algorithms after each student's first response to a problem in order to re-calculate the probability that the student knew the skill before the response. Then the algorithm accounts for the possibility that the student learned the skill during the problem in order to compute the probability the student will know the skill after the problem [6]. With the data from the logs, BKT parameters were fit by employing brute-force grid search [cf. 2].

After obtaining the assessments of student affect and prior knowledge at each problem, we assessed the relationship between the two. The following section shows both qualitative and quantitative estimates of these relationships for each affective state. Since our models provide confidences in their predictions as well as overall predictions, we conduct analyses using the confidences of the affect predictions rather than the proportion of binary predictions.

3 Studying the Relationship between Affect and Knowledge

3.1 Data Set

The detectors of student affect and student knowledge were applied to a data set consisting of five years of student usage of the ASSISTment system by four schools in New England, from 2004-2005 to 2008-2009. These four schools represent a diverse sample of students in terms of ethnicity and socio-economic status. Two districts were urban with many students requiring free or reduced-price lunches due to poverty, relatively low scores on state standardized examinations, and many students learning English as a second language. The other two districts were suburban, serving relatively wealthier populations. The affect models were applied to this much larger dataset. This data set included 8,454 students and a total of 1,568,974 student actions within the learning software.

3.2 Boredom and Student Knowledge

Boredom is less common when student skill is higher, as shown in Figure 2. This finding contrasts with predictions by Csikszentmihalyi [8] and Shernoff et al. [28], which would suggest that boredom should mostly occur when material is too easy relative to student skill. The linear trend is fairly modest (a difference of 5% in average boredom between material where the student has a high probability of knowing

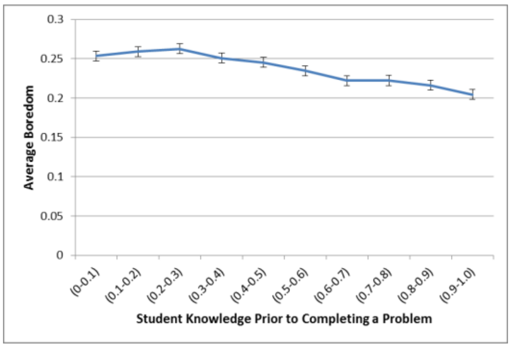


Fig. 2. The relationship between boredom and the probability that the student knows the skill. Note that the X axis denotes difficulty for the current problem for the current student, prior to the student completing the problems; i.e., the contextually hardest problems are on the left, and the contextually easiest problems are on the right.

the skill and material where the student has a very low probability of knowing the skill). However, due to the large sample size, the negative linear trend is statistically significant ($r = -0.157$, $F(1, 1560519) = 14223.174$, $p < 0.0001$). Note that a student term was included in the model (and all the statistical tests in this paper) to avoid violation of statistical independence.

3.3 Frustration and Student Knowledge

The relationship between frustration and student skill, shown in Figure 3, appears non-linear. Frustration appears to be significantly more common for students with very low skill and for students with very high skill, than for other students. When we fit a linear curve, there is a significant but small correlation between frustration and prior knowledge ($r = 0.093$, $F(1, 1560519) = 11647$, $p < 0.0001$). A parabolic curve (i.e., Frustration = (Knowledge – Mean(Knowledge))^2) achieves better fit ($r = 0.222$, $F(1, 1560519) = 63989$, $p < 0.0001$). The difference in BiC' values between these two models is 65,667, indicating that the parabolic curve fits the data substantially better

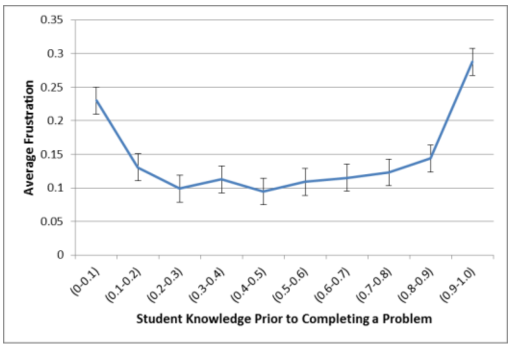


Fig. 3. The relationship between frustration and the probability that the student knows the skill.

than the linear function (differences in BiC' of ten or greater indicate substantial differences between models). The relationship between low skill and frustration accords with Flow theory, but the relationship between high skill and frustration is surprising, indicating that students may become frustrated when repeatedly given easy items.

3.4 Engaged Concentration and Student Knowledge

The incidence of engaged concentration is higher for more skilled students, as shown in Figure 4. The linear trend is fairly modest (a difference of 6% in average engaged concentration between material where the student has a high probability of knowing the skill and material where the student has a very low probability of knowing the skill). However, due to the large sample size, the linear trend is statistically significant ($r = 0.184$, $F(1, 1560519) = 13660.477$, $p < 0.0001$). In accordance with past studies [3, 24], engaged concentration is the most common affect when using ASSISTments regardless of student skill level.

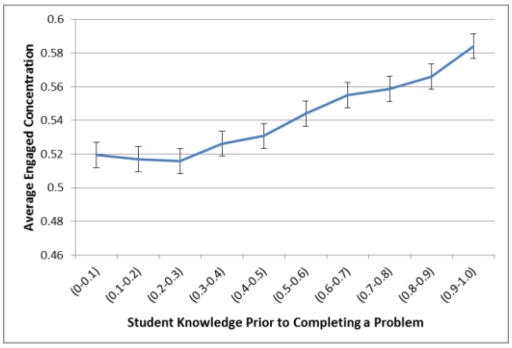


Fig. 4. The relationship between engaged concentration and the probability that the student knows the skill.

4 Discussion and Conclusion

Flow theory has emphasized the importance of achieving a balance between perceived challenge of a task and perceived skill for that task, to produce optimal student engagement (i.e., flow). In these models, an imbalance between challenge and skill would result in either boredom or frustration (or anxiety, which is not studied here).

In this paper, we study the relationship between these student affect and student knowledge within the context of an ITS, towards providing a concrete test of one aspect of Flow theory. We do so by applying automated detectors of student affect and knowledge to data from the ASSISTment system, a widely used intelligent tutoring system for middle school mathematics. By integrating these two types of detectors, we can analyze the frequency of each affective state for students with different levels of knowledge.

A limitation in this paper is that the model used for difficulty measures looked at estimations of actual knowledge and difficulty rather than a student's self-perceptions (as in from Flow theory). A challenge in obtaining measures of self-perception is that they may change the student's emotions and learning if obtained in real-time, and may be prone to memory limitations if obtained retrospectively. They also present some risk of demand effects. However, replicating this research with self-report measures would be a valuable step for future work.

Overall, we find that engaged concentration is the most likely affect, regardless of difficulty. This result shows that completing problems in ASSISTments is generally engaging, even when the problems are too easy or too difficult. Beyond this, problems are seen to become more engaging as student mastery increases, which contrasts somewhat with predictions made in Flow theory, which would predict that engagement would be reduced for the most challenging problems. (However, this result replicates a result seen in [17]). Flow theory predicts that these highly challenging problems will result in student frustration. Indeed, higher frustration is seen for the most challenging problems. However, higher boredom is also seen for these highly challenging problems, contrary to Flow theory. Boredom is generally lower for easy problems than hard problems, also contrary to Flow theory. In addition, higher frustration is seen for easy problems than for problems of middling difficulty, a finding that cannot be easily explained with Flow theory.

Given that these results are different from earlier predictions, it is worth thinking about their interpretation. There have been reports of boredom being associated with poorer learning [7, 19] and with disengaged behaviors that in turn lead to poorer learning [3]. Recent studies using other methods have also found that students become bored and disengaged when they find items difficult [1, 19]. These results accord with our findings that boredom is characteristic of less successful students rather than highly successful students. Perhaps these students are bored because they have given up on succeeding with the material, but must continue to work with the software. It may be that this type of boredom is more common in intelligent tutoring systems than boredom resulting from overly low challenge – especially since many tutors such as ASSISTments are designed to advance students when they reach mastery.

One possibility is that the relatively low boredom seen for easy items and the unexpected frustration seen on these items is due to the student's lack of control over problem difficulty. Perhaps a student who wishes to receive more challenging problems, but cannot obtain these problems within the software, becomes frustrated and upset with the software. In general, further research may be necessary in order to understand why students become frustrated with easy material. One possible approach would be to pop-up an automated question in this situation (detected frustration on easy material), asking students if they are frustrated and why. An interesting aspect of the current finding on frustration and student knowledge is that this result provides an account for a surprising result from previous studies. Past research has failed to find significant relationships between frustration and learning outcomes [cf.7, 22], contrary to theoretical predictions [13]. If unsuccessful students are not more likely to become frustrated, one would not expect to see such a relationship. In general, frustration appears to be a more complex construct than originally thought [cf. 13].

Overall, our findings suggest that there may be substantial holes in our understanding of the situations where different affective states emerge, during human-computer interaction. Current theory does not explain these results, and makes predictions that are in some cases contrary to the findings presented here. It is important to note that these findings only involve one intelligent tutor, and rely upon imperfect detectors of both affect and knowledge (though each of these detectors is approximately as good as the current state-of-the-art for sensor-free detection of these constructs). Replicating these results (or failing to) in other learning software will be an important step towards understanding the generality of these findings, and towards creating general principles for how intelligent tutoring systems should respond to users when they demonstrate these affective states. It is likely that we will find that each of the affective states can emerge in multiple situations, driven by differences in tutor design, and perhaps by individual differences as well. Hence, further investigation of the contexts of affect will be needed to fully understand these relationships.

Acknowledgements. This research was supported by grants NSF #DRL-1031398, NSF #SBE-0836012, and grant #OPP1048577 from the Bill & Melinda Gates Foundation. We also thank Zak Rogoff, Adam Nakama, Aatish Salvi, Adam Goldstein, and Sue Donas for their assistance in conducting the study.

References

1. Acee, T. W., Kim, H., Kim, H. J., Kim, J., Hsiang-Ning, R. C., and Kim, M. Academic boredom in under- and overchallenging situations. Contemporary Educational Psychology, 35 (2010), 17–27.
2. Baker R.S.J.d., Corbett A.T., Gowda S.M., Wagner A.Z., MacLaren B.M., Kauffman L.R., Mitchell A.P. and Giguere S. Contextual slip and prediction of student performance after use of an intelligent tutor. In Proc. UMAP 2010, 52-63.
3. Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., and 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. Int'l. J. Human-Computer Studies 68, 4 (2010), 223-241.
4. Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., and Rossi, L. Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. In Proc. EDM 2012, 126-133.
5. Conati, C., and Maclaren, H. Empirically building and evaluating a probabilistic model of user affect. User Modeling and User-Adapted Interaction 19, 3 (2009), 267-303.
6. Corbett, A.T., Anderson, J.R. Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction 4, 4 (1995), 253-278.
7. Craig, S. D., Graesser, A. C., Sullins, J., and Gholson, B. Affect and learning: an exploratory look into the role of affect in learning. J of Educational Media 29 (2004), 241–250.
8. Csikszentmihalyi, M. Flow:The Psychology of Optimal Experience. Harper-Row 1990.
9. Csikszentmihalyi, M., and Csikszentmihalyi, I.S. Optimal Experience: Psychological Studies of Flow in Consciousness. Cambridge, Cambridge University, 1988.

10. D'Mello, S. K., Craig, S. K., Gholson, B., Franklin, S., Picard, R. W., and Graesser, A. C. Integrating affect sensors in an intelligent tutoring system. In Proc. Intelligent User Interface 2005, AMC Press (2005).
11. D'Mello, S.K., Craig, S.D., Witherspoon, A. W., McDaniel, B. T., and Graesser, A. C. Automatic Detection of Learner's Affect from Conversational Cues. User Modeling and User-Adapted Interaction 18, 1-2 (2008), 45-80.
12. Dragon, T., Arroyo, I., Woolf, B.P., Burleson, W., Kaliouby, R.e., Eydgahi, H. Viewing Student Affect and Learning through Classroom Observation and Physical Sensors. In Proceedings of the International Conference on Intelligent Tutoring Systems 2008, 29-39.
13. Kort, B., Reilly, R., Picard, R. An Affective Model of Interplay between Emotions And Learning: Reengineering Educa-tional Pedagogy—Building A Learning Companion. In Proc. IEEE Int'l. Conf. on Advanced Learning Technology: Issues, Achievements and Challenges 2001, 43-48.
14. Lehman, B., D'Mello, S. K., and Person, N. All Alone with your Emotions: An Analysis of Student Emotions during Effortful Problem Solving Activities. Workshop on Emotional and Cognitive issues in ITS, Int'l Conf. on Intelligent Tutoring Systems (2008).
15. Miserandino, M. Children Who Do Well In School: Individual Differences in Perceived Competence And Autonomy In Above-Average Children. J.Ed Psych 88, (1996), 203-214.
16. 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 Proc. Learning Analytics and Knowledge (in press).
17. Pavlik, P. I., Jr. The microeconomics of learning: Optimizing paired-associate memory. Doctoral Dissertation, Carnegie Mellon University, 2005.
18. Pavlik Jr., P. I., Presson, N., Dozzi, G., Wu, S.-m., MacWhinney, B., and Koedinger, K. R. The FaCT (Fact and Concept Training) System: A new tool linking cognitive science with educators. In Proc. Cognitive Science Society 2007, 397-402.
19. Pekrun, R., Goetz, T., Daniels, L.M., Stupnisky, R.H., and Perry, R.P. Boredom in Achievement Settings: Exploring Control-Value Antecedents and Performance Outcomes of a Neglected Emotion. J. Educational Psychology 102, 3 (2010), 531-549.
20. Pilke, E.M. Flow Experiences in Information Technology Use. Int'l. J. Human-Computer Studies 61, 3 (2004), 347-357.
21. Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N.T., et al. The Assistment project: Blending assessment and assisting. In Proc. AIED 2005, 555-562.
22. Rodrigo, M.M.T. and Baker, R.S.J.d. Coarse-Grained Detection of Student Frustration in an Introductory Programming Course. In Proc ACM ICER 2009.
23. Rodrigo, M.M.T., Baker, R.S., Jadud, M.C., et al. Affective and Behavioral Predictors of Novice Programmer Achievement. In Proc. ACM-SIGCSE 2009, 156-160.
24. Sabourin, J., Mott, B., and Lester, J. Modeling Learner Affect with Theoretically Grounded Dynamic Bayesian Networks. In Proc. ACII 2011, 286-295.
25. Sedighian, K. Challenge-driven learning: A model for children's multimedia mathematics learning environments. In Proc. ED-MEDIA 1997.
26. Sessink, O., Beeftink, H., Tramper, J., and Hartog, R. Proteus: A Lecturer-Friendly Adaptive Tutoring System. J. Interac-tive Learning Research 18, 4 (2007), 533-554.
27. Shernoff, D. J., and Csikszentmihalyi, M. Flow in schools: Cultivating engaged learners and optimal learning environ-ments. In Handbook of Positive Psychology in Schools, R. Gilman, E. S. Huebner, & M. Furlong, Eds. Routledge, New York, 2009, 131-145.
28. Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., and Shernoff, E. S. Student engagement in high school classrooms from the perspective of flow theory. School Psychology Quarterly 18, 2 (2003), 158.