Which is more responsible for boredom in intelligent tutoring systems: students (trait) or problems (state)?

William Hawkins, Neil Heffernan
Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA
{bhawk90, nth}@wpi.edu

Ryan S.J.d. Baker
Department of Human Development
Teachers College Columbia University
New York, NY
baker2@exchange.tc.columbia.edu

Abstract— **Boredom is unpleasant, and has been repeatedly shown to be associated with poor performance and long-term disengagement in educational contexts. Boredom is prevalent within a range of online learning environments, has been shown to correlate negatively with learning in those environments, and often precedes disengaged behaviors such as off-task behavior and gaming the system. Therefore, it is important to identify the causes of boredom in these environments. In psychology research, there is ongoing debate about the degree to which individual students are prone to boredom ("trait" explanations) or the degree to which boredom is driven by state-based factors, such as the design of the learning environment. In this study, we apply an unobtrusive computational detector of student boredom to log data from an intelligent tutoring system to determine whether state or trait factors better predict the prevalence of boredom in students using that system. Knowing which type of factor better predicts boredom in a specific system can help us to narrow down further research on why boredom occurs and what steps should be taken to mitigate boredom's negative effects.**

Keywords—intelligent tutoring system; boredom; "state" vs. "trait"

I. Introduction

Over the past several decades, there has been a considerable degree of interest regarding student boredom during learning. Many agree that boredom is an unpleasant or negative experience [19, 25] but propose different potential causes and effects of the emotion, and disagree about its impact and how to respond to it [6].

According to Belton and Priyadharshini's survey of boredom research [6], most psychological research concerned with the causes of boredom has posited on a dichotomy of two possible causes that has variously been referred to as "responsive" vs. "chronic," "agitated" vs. "apathetic," "dispositional" vs. "situational," and most commonly "state" vs. "trait".

The "state" construct of boredom describes boredom as being caused by a specific situation or experience that is objectively boring, where external stimulus is lacking [6]. A specific situation or experience can lack stimulation for a variety of reasons [cf. 30, 21, 6, 19, 25, 22].

Alternatively, state boredom may be caused by temporary aspects of the student, for example fatigue. In a 2000 study conducted with 170 American university students, 17% identified fatigue as an indicator that they were bored, 8% identified it as a cause of boredom, and 15% reported that sleeping was one way they coped with boredom [19]. Therefore, the causes of state boredom can be thought of in terms of the current state of the person, as well as the nature of the activity.

Another viewpoint is that boredom is caused by student traits [6], where certain individuals are more prone to boredom than others. Various studies have used the Boredom Proneness Scale [13], a questionnaire measure, to assess the susceptibility of different individuals to boredom, and have found links between this boredom "trait" and other personality characteristics [13], as well as to various negative and destructive behaviors [19, 30].

Boredom in education has been studied in terms of state and trait constructs [21], as it has in other contexts such as in the workplace [6]. Within education, state explanations for boredom have blamed schools and the design of classroom activities, arguing that boredom is caused by meaningless or repetitive tasks [27], overly abstract activities [8], and tasks being too challenging [9] or not challenging enough [23]. However, Larson and Richards also found that some students associated boredom with fatigue, and that boredom co-occurred with tiredness and drowsiness [21].

On the trait side of the debate, the dispositions that students bring to school have been blamed for the boredom they experience there [17, 14]. Additionally, learning goals, perceived level of control, and the relative value a student places in a skill or activity have all been argued to be associated with boredom in educational settings [25].

To determine the relative effects of state and trait boredom among middle school students, Larson and Richards [21] measured boredom experienced by students over the course of a week using randomly timed surveys. They found that boredom was prevalent both in and out of school, and that it depended more on the individual student than on the subject or the activity. They found that boredom students experienced in and out of school was highly correlated (r = 0.68). However, both subject and activity also had a substantial influence on boredom in this research.

Understanding why students become bored is important, as boredom has been shown to be associated with poorer learning

This research was supported by a GAANN fellowship, a CAREER grant, National Science Foundation award #DRL-1031398 and grant #OPP1048577 from the Bill & Melinda Gates Foundation.

[25, 13] as well as long-term disengagement [14]. Within intelligent tutoring systems (ITSs), it has been shown that boredom is one of the most persistent affective-cognitive states [4], and that it leads to gaming the system, or "attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material" [4]. Gaming the system has also been linked to poorer learning, both in the short-term [3, 18] and in the long-term [3].

Boredom has also been shown to lead to off-task behavior in ITSs [5], which is also associated with poorer learning [20] and long-term disengagement [16]. However, Baker et al.'s research [5] suggested that off-task behavior can relieve boredom and allow the student to refocus on their work with a lower probability of experiencing boredom later. Regardless, both gaming the system and off-task behavior take up considerable time [3], giving students less time to use ITSs constructively and learn from them. Given the increasing use of ITSs within education, the effects boredom has within ITSs and on learning in general, it is important to study the causes of boredom in ITSs.

In order to study boredom adequately in ITSs, it is important to have a broadly applicable method of assessing the presence and intensity of boredom. While many studies and systems have used physical sensors to detect boredom [1, 11], these approaches can be difficult to scale within large numbers of classrooms due to issues such as internet bandwidth, cost of sensors, and breakage under classroom conditions.

Due to the restricted applicability of models built using physical sensors, researchers have recently worked to develop affect detectors based solely on log files for various platforms, including AutoTutor [10], Prime Climb [7], Crystal Island [28], and ASSISTments [29]. As these types of detectors rely only on log data, they can be applied to large amounts of data, allowing datasets to be labeled for use in exploratory analyses.

Given the need to study boredom in ITSs and the availability of broadly applicable log-based boredom detectors, this work uses results from a real-time log-based boredom detector [29] to determine whether boredom in ITSs is caused more by state or trait characteristics. Specifically, linear regression models are constructed to determine whether state or trait factors better predict boredom, as done previously by others for studying whether gaming the system is better predicted by state or trait factors [2, 24].

In addition to state vs. trait analysis, models incorporating proxies for fatigue are fit. Due to previously reported associations between fatigue and boredom [21, 19] and the finding that off-task behavior in ITSs helps relieve boredom [5], we hypothesize that measures of boredom and fatigue within this study will be positively correlated.

Once it is determined which construct better predicts boredom in ITSs, it can be studied closer to determine what steps should be taken to respond to it. If boredom is better predicted as a state variable, then the content and interface of the system should be studied further and improved to reduce boredom. If on the other hand boredom is better predicted as a trait variable, then student characteristics should be studied further to determine how to change the system specifically for students prone to boredom. By narrowing down the type of approaches that are likely to address boredom, we can move towards reducing its prevalence and impacts in ITSs, with the goal of improving both student engagement and learning.

Section II introduces the dataset and methods used in this work. Section III presents the results of the analyses, and Section IV concludes with discussion and possible directions for future work.

II. Methods: Overview

A. Tutor and Data

For this work, data from the ASSISTments intelligent tutoring system [15] was used. ASSISTments is a web-based ITS used primarily by middle- and high-school students. In the 2012-2013 school year, it is being used by 40,000 students, mostly in the Northeastern USA, around once a week. While using ASSISTments, students are assessed based on their performance within the system, which is reported back to teachers. Additionally, students are assisted while working through problem sets in three main ways: hint messages, which progress from high-level hints to a "bottom-out hint" containing the answer to the problem; feedback given when the student gives incorrect answers; and scaffolding, where the system breaks a problem down into sub-problems. An example of a student working through a problem in ASSISTments is shown in Fig. 1.

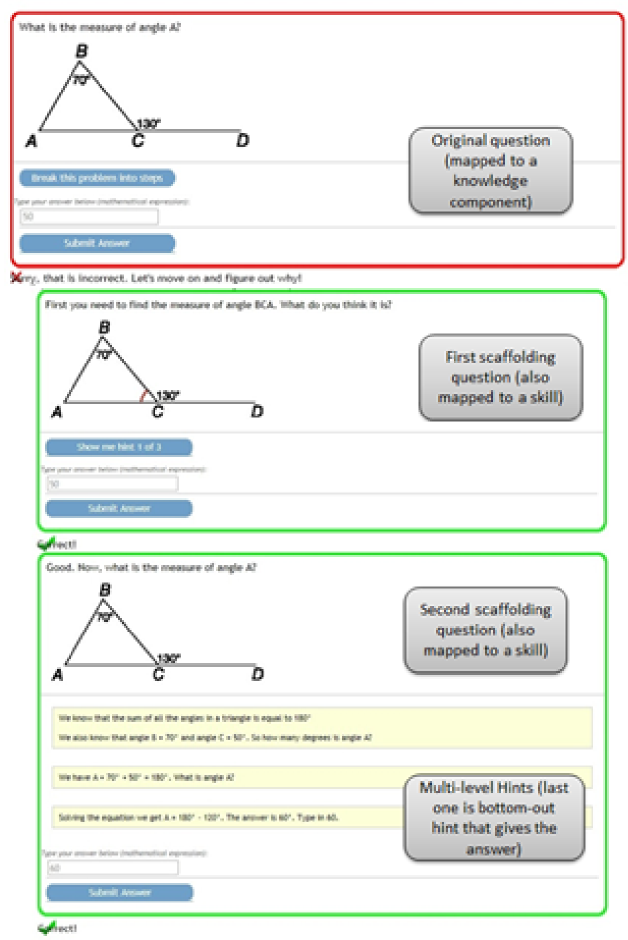

Fig. 1. Example of a student working through a problem in ASSISTments, from top to bottom. The student first answers the question incorrectly, resulting in feedback. The problem is then scaffolded, and the student answers the first scaffolding question correctly. Finally, the student clicks through all the hint messages of the second scaffolding question, reaching the "bottom-out hint," which contains the answer the student types in to solve the problem.

Within this study, data previously collected was used in analysis. The data was collected from ASSISTments data logs from September 2004 to May 2005 for 724 students from four central Massachusetts middle schools, consisting of 107,382 problem attempts. This data set was chosen because the affect detector had already been applied to it and it had been used in a previous study [e.g., 29]. Each problem attempt includes the ID of the student that made the attempt, the problem ID, the relevant "skill" being tested by the problem (e.g., multiplication, area, equation-solving, etc.), and the "type" of problem or method of answering the question (e.g., multiple choice, fill-in, etc.). There were 10 different types of problems and 70 skills represented.

Additionally, each problem attempt was labeled with a real-valued confidence level between 0 and 1 that boredom was present. Confidences can be interpreted as the detector's estimate of the probability that the student was bored at a specific time. These values were computed using a real-time boredom detector, discussed in full detail in [29]; in brief, this detector was developed by synchronizing thousands of field observations of boredom (with inter-rater reliability over 0.6) with log files, and using data mining to infer the human codes. The mean boredom confidence across all problem attempts was 0.2469 (SD = 0.1293). Confidences were used instead of binary predictions of boredom, in order to leverage the detector's ability to distinguish cases it is unsure of – for instance, claiming that a case with 51% certainty is identical to a case with 100% certainty, but is fundamentally different from a case with 49% certainty, throws out considerable information and increases the noise in the data set.

B. Modeling Method

To determine whether boredom can be modeled better as a state or trait construct, we fit linear regression models to the data to predict the confidence of the boredom detector. One model is trained using only the problem ID from each problem attempt as a predictor, while the other uses only the student ID from each problem attempt as a predictor. The state theory hypothesizes that the difference between problems will predict much of the variance in student boredom, while the trait theory hypothesizes that the difference between students will predict much of the variance in boredom. The R^2 and Bayesian Information Criterion (BiC'; [26]) values of the predictions made by these models can then be used to assess which construct is a better predictor of boredom. Additionally, a third model is fit with both problem ID and student ID as predictors.

Similar procedures to that described above have been performed for gaming the system in order to assess whether it is better viewed as a state or trait construct [2, 18, 24]. Baker's analysis found that lessons predicted gaming better than students, which was contradicted by the findings of the other two groups [18, 24]; further unpublished analysis conducted by two of these research groups working together suggests that this may be due to differences in the operational definition of gaming used in the different studies. In line with Baker's method, we use IDs rather than the average confidence of affect detector results; however, we analyze state-level prediction at the problem-level as in [24], rather than analyzing at a coarser grain-size.

Finally, a proxy for fatigue is computed and added to the data. In this study, the proxy for fatigue is operationalized as the number of minutes that passed since the student last took a break of a certain number of minutes. We hypothesize that fatigue will be a successful predictor due to previous findings of relationships between fatigue and boredom [19, 21] and the relationship between boredom and off-task behavior, in which off-task behavior appears to relieve boredom [5]. We hypothesize that the longer a student goes without an opportunity to relieve their boredom, the higher their boredom will be. A number of linear regression models are built using only this fatigue statistic for different time durations that constitute a break. The best of these fatigue attributes is then combined with predictors from the other models described above and tested.

III. ANALYSES

A. State vs. Trait

The first research goal was to determine whether state (problems) or trait (students) was a better predictor of boredom. For this analysis, two linear regression models were fit: one that used only problem ID as a predictor, and one that only used student ID. The target attribute for both was the detector's real-valued confidence that boredom was present. These models were evaluated using two measures: R^2, and the Bayesian Information Criterion (BiC'; [26]), which calculates the degree to which a model's predictions are better than what would be expected solely from the number of parameters used. Lower values of BiC' are better, and a difference of six or more between the BiC' values of two models is considered equivalent to being statistically significant at the $p < 0.05$ level [26]. The R^2 values were computed using five-fold cross-validation, where the folds were stratified both by problem ID and student ID. The same folds were used for all models, which were built using Matlab's LinearModel class. The BiC' values were computed over the entire dataset. The results are shown in Table I.

TABLE I. MODEL FITTING RESULTS FOR STATE VS. TRAIT. R^2 CALCULATED USING FIVE-FOLD CROSS-VALIDATION, BIC' CALCULATED OVER THE ENTIRE DATASET

Model	R^2	BiC'
Baseline	0.0000	0.00
Problem ID	0.0516	-13,002
Student ID	0.0061	2,845
Both	0.0818	-11,012

As Table I shows, the model based only on problem IDs is significantly more predictive than that based only on student IDs, suggesting that the incidence of boredom is more dependent on the problem being attempted rather than on the student attempting it. Additionally, the student model does worse than the Baseline model (which predicted the mode of boredom confidences, 0.1273, for all problem attempts) judging by its large positive BiC' value. A third model that used both student IDs and problem IDs as predictors achieved a higher R^2 value, but did worse for the number of parameters it used compared with the model that only used problem IDs, based on their respective BiC' values.

B. Fatigue

When predicting boredom, it may be helpful to include other factors, such as the current state of the student. One key aspect of the student is whether the student is fatigued. To test this hypothesis, a proxy for the construct of fatigue was added to each problem attempt in the dataset. Fatigue was calculated in two ways, producing models we refer to as MFatigue and PFatigue.

MFatigue was calculated as the number of minutes that had passed since the student had taken a break of a certain number of minutes. For example, the attribute "MFatigue(60)" (M stands for minutes) is defined as the number of minutes that have elapsed since the student in question last took a break of 60 minutes or more.

PFatigue was calculated as the number of problems completed by the student since last having a break of a certain number of minutes. We hypothesized that the monotony, and therefore boredom, that the students experienced would increase with the number of problems they completed. For example, the attribute "PFatigue (60)" is defined as the number of problems a given student has completed since last having a break of 60 minutes or more.

A number of linear regression models were built using each fatigue statistic for different amounts of minutes that constituted a break. This was done since it was uncertain how long of a break was necessary to "reset" a student's level of fatigue. The results were computed using five-fold cross-validation with the same five folds used for the state and trait models. The results are shown in Table II.

TABLE II. MODEL FITTING RESULTS FOR FATIGUE. MODELS INCLUDING BOTH FATIGUE AND OTHER FACTORS GIVEN AT BOTTOM

Model	R^2	BiC'
MFatigue(60)	0.0000	-354
MFatigue(30)	0.0000	-537
MFatigue(15)	0.0000	-938
MFatigue(10)	0.0002	-1,469
MFatigue(5)	0.0006	-2,136
PFatigue(60)	0.0013	-2,784
PFatigue(30)	0.0016	-3,093
PFatigue(15)	0.0029	-3,914
PFatigue(10)	0.0039	-4,495
PFatigue(5)	0.0038	-4,473
Problem ID, PFatigue(10)	0.0644	-15,836
Problem ID, Student ID, PFatigue(10)	0.0878	-12,229

Both the problem-based and the minute-based fatigue models were better than the baseline model and the student/trait model. However, the problem-based fatigue models were much more predictive than their minute-based counterparts. Using 5 or 10 minutes as the duration of break needed to reset the student's fatigue produced the best models for predicting boredom from fatigue. The best combination of problem IDs with a fatigue attribute used a break of 10 minutes for resetting fatigue, and significantly improved upon the model that only predicted boredom from the problem ID. This combined (problem ID, fatigue) model was the most predictive model found in these analyses, with a difference in BiC' of almost 3,000 from the next best model (problem ID). Similarly, adding PFatigue(10) to the (problem ID, student ID) model also improved performance, though this model's performance still fell short of the problem ID model (or the problem ID, PFatigue(10) model).

Having found that our hypothesis about fatigue being a strong predictor of boredom was correct, we next set out to determine if our hypothesis about higher fatigue leading to higher boredom was correct. However, it turned out that "fatigue" was instead negatively correlated with boredom; e.g., the longer it had been since the student took a break, the less bored they were. Examining the linear regression models based on fatigue revealed that fatigue had a negative coefficient, and correlation analysis showed that all calculations of fatigue were negatively correlated with boredom. The strongest of these correlations was for PFatigue(10), where $r = -0.20$. It should be noted this correlation was calculated over the entire dataset, and therefore does not correspond to the cross-validated R^2 value reported for the PFatigue(10) linear regression model in Table II.

C. Type and Skill

Since problems were found to be more predictive than students, we focused on state explanations of boredom for the remainder of our analyses. Additionally, it has been found that boredom is associated with the difficulty of tasks within achievement settings [25, 29]. Therefore, as a preliminary analysis of which specific problem features cause boredom in ASSISTments, the association between the type and skill of problems, as defined above, and the level of boredom experienced on them was studied.

Linear regression models were tested using type and skill, using five-fold cross-validation and the same folds used for all previous models. As was done for problems and students, separate models were fit for type and skill, and a third model combining the two was also fit. The results are shown in Table III.

TABLE III. MODEL FITTING RESULTS FOR TYPE AND SKILL

Model	R^2	BiC'
Type	0.0039	-4,290
Skill	0.0131	-7,063
Type and Skill	0.0184	-8,530
Type, Skill, PFatigue(10)	0.0294	-11,491

Both the skill and type models outperformed the student ID model, but performed more poorly than the problem ID model. In general, the results also indicate that the skill of a problem is more predictive than its type. Combining type and skill into one model further improved performance, but still not to the same level as the problem ID model. Additionally, PFatigue(10) proved useful again as adding it to the (type, skill) model significantly improved performance – but again, still not to the level of the problem model. Therefore, skill and type seem important, but it appears there are other important problem features missing from this analysis due to the differences in predictive power of the type and skill model and the problem ID model. A full study of which problem features lead to boredom [cf. 12] is warranted.

IV. Discussion and Future Work

Contrary to previous work [21], we found that state (individual problems) was more predictive of boredom than trait (individual students) in an intelligent tutoring system, ASSISTments. Two possible explanations for this are the different methodologies employed by the two studies, and the different contexts in which boredom was studied.

First, boredom in this work is measured by an unobtrusive computational detector, whereas the prior study measured it using questionnaires and interviews [21]; these differences in measurement may change the results in multiple ways. For instance, potential differences in student comfort and attitudes in self-reporting their boredom could drive the appearance of a student-level effect within self-report methodologies. Second, this work studied boredom at the individual problem level within an intelligent tutoring system, whereas the prior study looked at boredom at a much higher level. The prior study looked at variations in boredom across different subjects like mathematics and English; across different activities within the school environment (such as listening to a teacher or student, reading, and taking a test or quiz); and across school, home, and public environments [21].

Since it appears that boredom is caused more by individual problems than by students, at least in the context of ASSISTments, future work should focus on identifying which features of problems are the most responsible for boredom. Identifying such features will help inform the design of problems in the future to reduce boredom, increase learning rates, and reduce long-term disengagement.

In this work, two features of problems were considered: the skill (multiplication, equation-solving, etc.) and the type of problems (multiple choice, fill-in, etc.). The linear regression model that considers these features together achieves a paltry R^2 of 0.0184, leaving a significant portion of the R^2 achieved by the problem ID model (0.0516) unexplained. Future work should explore what other properties of problems contribute to boredom.

The level of knowledge a student possesses in a given skill has been shown to explain some of the variance in boredom, though one study found boredom to be higher among highly skilled students [21] while a set of five other studies have found the opposite relationship [25]. Therefore, the relative difficulty of a problem or skill may help predict boredom in an intelligent tutoring system context. Looking beyond the aforementioned high-level problem features, a process for determining relevant features similar to what was done by Doddannara et al. [12] may be relevant and useful for following up the results seen here.

Additionally, we found that modeling a proxy for student “fatigue” can add predictive power to the problem model. The best operationalization of fatigue was the number of problems completed since the student last had a break of 10 minutes or more. However, somewhat surprisingly, this measure we constructed as a proxy for fatigue was negatively correlated with boredom. From the results and above analysis, it appears there is merit in considering attributes like fatigue that are not specific to problems (i.e., not uniquely identified by problem ID), but that describe the “session” (what a student completes in a single sequence of activity) or the student’s current state. It may be worth considering further attributes of this nature, such as how many times the student has seen the current skill, or how the current problem relates to previous problems in the problem set in terms of their other attributes, such as difficulty, skill or type.

In this work, we have shown that (at least within the ASSISTments ITS) boredom is better explained by problems (state) than students (trait). We did initial research into the specific components of state boredom within an ITS by fitting models to attributes of problems (type, skill) and the session (fatigue), both of which performed better than the baseline model at predicting boredom. Future work should consider additional factors such as problem difficulty, student knowledge (which is specific to individual students but changes dramatically over time, and which is more consistent with state than trait), and the amount of previous practice the student has had, among other state features.

Another valuable area of work is to follow up this work with replications in other online learning environments, and using alternate operationalizations of boredom (as in the follow-ups to [2] by Gong et al. [18] and Muldner et al. [24] for gaming). By better understanding the factors that influence whether a student becomes bored while using an ITS, we may be able to develop more emotionally-sensitive learning systems that lead to better learning and higher long-term engagement.

Acknowledgments

We would like to acknowledge the many funders of ASSISTments Platform found here: http://www.webcitation.org/5ym157Yfr.

We would also like to acknowledge Maria O.C.Z. San Pedro for facilitating our use of the data set, and Lisa Rossi for her help with proofreading.

All of the opinions expressed in this paper are those solely of the authors and not those of our funding organizations.

References

[1] I. Arroyo, B.P. Woolf, D. Cooper, W. Burleson, K. Muldner, R. Christopherson, “Emotion Sensors Go To School,” in *Proc. 14th Int. Conf. on Artificial Intelligence In Education*, Brighton, UK, 2009, pp. 17-24.

[2] R.S.J.d. Baker, “Is Gaming the System State-or-Trait? Educational Data Mining Through the Multi-Contextual Application of a Validated Behavioral Model,” in *Proc. Workshop on Data Mining for User*

Modeling at the 11th Int. Conf. on User Modeling, Corfu, Greece, 2007, pp. 76-80.

[3] R.S.J.d. Baker, A.T. Corbett, K.R. Koedinger, A.Z. Wagner, "Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System"," in *Proc. ACM CHI 2004: Computer-Human Interaction*, Vienna, Austria, 2004, pp. 383-390.

[4] R.S.J.d. Baker, S.K. D'Mello, M.M.T. Rodrigo, A.C. Graesser, "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*, vol. 68 no. 4, pp. 223-241, Apr. 2010.

[5] R.S.J.d. Baker, G. Moore, A. Wagner, J. Kalka, M. Karabinos, C. Ashe, D. Yaron, "The Dynamics Between Student Affect and Behavior Occuring Outside of Educational Software," *in Proc. 4th bi-annual Int. Conf. on Affective Computing and Intelligent Interaction*, Memphis, TN, 2011.

[6] T. Belton, E. Priyadharshini, "Boredom and schooling: A cross-disciplinary exploration," *Cambridge Journal of Education,* vol. 37, no. 4, pp. 579−595, Dec. 2007.

[7] C. Conati, H. Maclaren, "Empirically building and evaluating a probabilistic model of user affect," *User Modeling and User-Adapted Interaction*, vol. 19, no. 3, pp. 267-303, Aug. 2009.

[8] J. Condry, "The role of incentives in socialization," in *The hidden costs of reward: new perspectives on the psychology of human motivation*, M. Lepper and D. Greene, Eds. New York: John Wiley & Sons, 1978, pp. 179–192.

[9] C. Cullingford, *The best years of their lives? Pupils' experience of school.* London: Kogan Page, 2002.

[10] S.K. D'Mello, S.D. Craig, A.W. Witherspoon, B.T. McDaniel, A.C. Graesser, "Automatic Detection of Learner's Affect from Conversational Cues," *User Modeling and User-Adapted Interaction*, vol. 18, no. 1-2, pp. 45-80, Feb. 2008.

[11] S.K. D'Mello, A.C. Graesser, "Multimodal semiautomated affect detection from conversational cues, gross body language, and facial features," *User Modeling and User-Adapted Interaction*, vol. 20, no. 2, pp. 147-187, Jun. 2010.

[12] L.S. Doddannara, S.M. Gowda, R.S.J.d Baker, S.M. Gowda, A.M.J.B. de Carvalho, "Exploring the relationships between design, students' affective states, and disengaged behaviors within an ITS," to appear in *Proc. 16th Int. Conf. on Artificial Intelligence in Education*, Memphis, TN, in press.

[13] R. Farmer, N. Sundberg, "Boredom proneness: the development and correlates of a new scale," *Journal of Personality Assessment*, vol. 50, no. 1, pp. 4–17, 1986.

[14] E. Farrell, "Giving Voice to High School Students: Pressure and Boredom, Ya Know What I'm Sayin'?" *American Educational Research Journal*, vol. 25, no. 4, pp. 489-502, Dec. 1988.

[15] M. Feng, N.T. Heffernan, K.R. Koedinger, "Addressing the assessment challenge in an Intelligent Tutoring System that tutors as it assesses," *User Modeling and User-Adapted Interaction*, vol. 19, no. 3, pp. 243-266, Aug. 2009.

[16] J.D. Finn, "Withdrawing From School," *Review of Educational Research*, vol. 59, no. 2, pp. 117-142, Summer 1989.

[17] T. Gjesme, "General Satisfaction and Boredom at School as a Function of the Pupils' Personality Characteristics," *Scandinavian Journal of Educational Research*, vol. 21, pp. 113-46, 1977.

[18] Y. Gong, J. Beck, N. Heffernan, E. Forbes-Summers, "The impact of gaming (?) on learning at the fine-grained level," in *Proc. 10th Int. Conf. on Intelligent Tutoring Systems (ITS2010) Part 1*, pp. 194-203, Pittsburgh, PA, 2010.

[19] M.B. Harris, "Correlates and Characteristics of Boredom Proneness and Boredom," *Journal of Applied Social Psychology*, vol. 30, no. 3, pp. 576–598, Mar. 2000.

[20] N. Karweit, R.E. Slavin, "Time-On-Task: Issues of Timing, Sampling, and Definition," *Journal of Educational Psychology*, vol. 74, no. 6, pp. 844-851, Dec. 1982.

[21] R.W. Larson, M.H. Richards, "Boredom in the Middle School Years: Blaming Schools versus Blaming Students," *American Journal of Education*, vol. 99, no. 4, pp. 418-443, Aug. 1991.

[22] W. Mikulas, "The essence of boredom," *Psychological Record*, vol. 43, no. 1, pp. 3–13, Winter 1993.

[23] G. Moneta, M. Csikszentmihalyi, "The effect of perceived challenges and skills on the quality of subjective experience," *Journal of Personality*, vol. 64, no. 2, pp. 275–310, Jun. 1996.

[24] K. Muldner, W. Burleson, B. Van de Sande, K. VanLehn, "An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts," *User Modeling and User-Adapted Interaction - Special Issue on Data Mining for Personalized Educational Systems*, vol. 21, no. 1-2, pp. 99-135, Apr. 2011.

[25] R. Pekrun, T. Goetz, L.M. Daniels, R.H. Stupnisky, R.P. Perry, "Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion," *Journal of Educational Psychology,* vol. 102, no. 3, pp. 531-549, Aug. 2010.

[26] A.E. Raftery, "Bayesian Model Selection in Social Research," *Sociological Methodology,* vol. 25, pp. 111-163, 1995.

[27] K. Reid, *Disaffection from school*. London: Methuen, 1986.

[28] J. Sabourin, J. Rowe, B. Mott, J. Lester, "When Off-Task is On-Task: The Affective Role of Off-Task Behavior in Narrative-Centered Learning Environments," in *Proc. 15th Int. Conf. on Artificial Intelligence in Education*, pp. 534-536, Auckland, New Zealand, 2011.

[29] M. San Pedro, R.S.J.d Baker, S. Gowda, N. Heffernan, "Towards an Understanding of Affect and Knowledge from Student Interaction with an Intelligent Tutoring System," to appear in *Proc. 16th Int. Conf.Artificial Intelligence in Education Conference*. Memphis, TN, in press. Available online at http://www.columbia.edu/~rsb2162/AIED_SBGH.pdf

[30] S. Vodanovich, "Psychometric measures of boredom: A review of the literature," *Journal of Psychology*, vol. 137, no. 6, pp. 569–595, 2003.