"Yes!": Using Tutor and Sensor Data to Predict Moments of Delight during Instructional Activities

Kasia Muldner, Winslow Burleson, and Kurt VanLehn

Arizona State University
{Katarzyna.Muldner,Winslow.Burleson,Kurt.VanLehn}@asu.edu

Abstract. A long standing challenge for intelligent tutoring system (ITS) designers and educators alike is how to encourage students to take pleasure and interest in learning activities. In this paper, we present findings from a user study involving students interacting with an ITS, focusing on when students express excitement, what we dub "yes!" moments. These findings include an empirically-based user model that relies on both interaction and physiological sensor features to predict "yes!" events; here we describe this model, its validation, and initial indicators of its importance for understanding and fostering student interest.

Keywords: interest, motivation, empirically-based model, sensing devices.

1 Introduction

In some cultures, the classic "yes!" gesture is to clench the fist of one's dominant arm, jerk the arm downward and exclaim "yes!" - everyone understands this as an expression of triumphal victory. When we noticed this behavior among students using our physics tutoring system, we began to wonder about it. For instance, what causes a "yes!" during tutoring? Is the "yes!" behavior a desirable outcome in itself or is it also associated with other desirable outcomes?

Because we are interested in building affective learning companions, we are also interested in how a companion could use students' "yes!" behavior for its own ends, such as increased bonding with the student. This requires, however, that the companion can detect "yes!" behaviors in real time. This paper reports our progress on addressing these issues and questions, including:

1. *Is the "yes!" behavior a desirable outcome for a tutoring system or associated with one?* We argue from the literature that it is both.
2. *What causes "yes!" events and how can we increase their frequency?* We compare "yes!" episodes with ones where a "yes!" could have occurred but did not. This descriptive analysis sets the stage for future work on what could cause an increase in "yes!" events.
3. *How can "yes!" events by used by tutors, learning companions or other agents?* We present a review of the literature that suggests some possibilities.

P. De Bra, A. Kobsa, and D. Chin (Eds.): UMAP 2010, LNCS 6075, pp. 159–170, 2010.

4. *Can a "yes!" event be detected more accurately than a baseline approach?* We developed a regression model based on sensor and tutor log data analysis that has high accuracy.

The rest of this introduction contains literature reviews that address points 1 and 3, and a review of related work on affect detection (point 4).

1.1 The Likely Role of "yes!" in Learning and Interest

As we describe in Sect. 3, we view "yes!" as a class of brief expressions of (possibly highly exuberant) positive affect. Positive affect has been linked to increased personal interest [1, 2], which is in turn associated with a facilitative effect on cognitive functioning [3], and improved performance on creative problem solving and other tasks [4], persevering in the face of failure, investing time when it is needed and engaging in mindful and creative processing (for a review see [5]). Although there is work in the psychology community on how interest develops and is maintained (e.g., [6, 7]), to date there does not yet exist sufficient work on these topics to understand the role of positive affect in general and of "yes!" events in particular, so calls for additional research are common (e.g., [8]).

We should point out, however, that while positive affect could itself be considered a desirable property during tutoring, it has not always shown strong correlations with learning [9]. For instance, doing unchallenging problems may make students happy but may not cause learning. However, the "yes!" expression of positive affect may well be correlated with learning, because as we show later, "yes!" occurs only after the student has been challenged, and challenge fosters learning [10].

1.2 How Can "yes!" Events Be Used during Tutoring and Learning?

In general, about 50% of human tutor interventions relate to student affect [11], highlighting the importance of addressing affect in pedagogical interactions. As far as addressing "yes!" events, work on the impact of tutorial feedback provides some direction regarding how "yes!" detection can be valuable to a tutoring system for generating subsequent responses. For instance, praise needs to be delivered at the right moment, e.g., be perceived as representative of effort and sincere, to be effective [12], and so a "yes!" event may be exactly the right time for an agent to give praise.

If "yes!" events do predict increased learning, interest and motivation, then they can be used as proximal rewards for reinforcement learning of agent policies. For instance, Min Chi et al. [13] found that a tutorial agent's policies could be learned given a distal reward, namely, a students' learning gains at the end of six hours of tutoring. It seems likely that even better policies could be learned if the rewards occurred more frequently. That is, if a "yes!" event occurs, then perhaps the most recent dialogue moves by the agent could be credited and reinforced.

1.3 Related Work on Detecting Brief Affective States

Affect recognition has been steadily gaining prominence in the user modeling community, motivated by the key role of affect in various interactions. Like us, some researchers have proposed models for identifying a single emotion. For instance, Kappor et al. [14] rely on a sensor framework , incorporating a mouse, posture chair, video camera and skin conductance bracelet, to recognize frustration. McQuiggan and Lester [15] describe a data-driven architecture called CARE for learning models of empathy from human social interactions. In contrast to Kappor’s and our work, CARE only uses situational data as predictors for empathy assessment. Others have focused on identifying a set of emotions. Cooper et al.’s [16] four linear regression models each predict an emotion (frustration, interest, excitement, confusion). Like our work, these models are built from a combination of tutor and sensor log data, although only we explore the utility of eye tracking and student reasoning data. D’Mello et al. [17] use dialog and posture features to identify four affective states (boredom, flow, confusion, and frustration). In Conati’s model, [18] a set of six emotions are assessed (joy/regret, admiration/reproach, pride/shame) from tutor log data, subsequently refined to include one sensor modality, namely an EEG [19].

While there is some work on modeling users with eye tracker information, most of it has focused on how attention shifts predict focus (e.g., [20]), or how pupillary response predicts cognitive load [21]. This latter work is inspired by findings in psychology showing that pupillary response is increased by cognitive load [22]; likewise, affect also increases pupillary size [23]. However, results from experiments less tightly controlled than traditional psychology ones have been mixed, with many failing to find the anticipated link between pupillary response and state of interest (e.g., [24]). In the past we investigated how *only* pupillary response distinguishes different types of affect [25], and did not propose a model based on our results. In contrast, here we present a model that relies on a broad range of features across both interaction and sensor data to predict “yes!” moments. In doing so, we provide insight into the utility of pupillary information for predicting “yes!” events.

In short, although others have investigated predicting positive affective states, including joy [26], engagement [17] and excitement [16], our work distinguishes itself in several ways. First, we identify a novel set of features unique to “yes!”, including time on task, degree of reasoning and pupillary response. A more important difference relates to our methodology. A fundamental challenge in inferring affect from data is finding the appropriate gold standard against which to compare a model’s predictions. A common approach is to elicit affect information by explicitly querying users [16, 26]. This approach has the potential to be disruptive, thus resulting in inaccurate affect labels; it can also miss salient moments of interest (i.e., when affect is actually occurring). Another common approach relies on using human coders to identify affect in users [17], a technique that also suffers from limitations since human coder performance can be variable [17]. In contrast, we rely on talk-aloud for obtaining affective labels. Doing so has the potential to avoid the above pitfalls, because it is a form of naturally occurring data that has been shown to not interfere with the task at hand [27]. Talk-aloud is also used in [28], although there, only conversational cues are considered as affect predictors, while we use an array of tutor and sensor features.

2 Obtaining Data on "yes!" Moments

We obtained data on "yes!" moments from a previous user study we conducted [25], which involved students interacting with an intelligent tutoring system (ITS) for introductory Newtonian physics. This ITS, referred to as the Example Analogy (EA)-Coach [29], provides support to students during problem solving in the presence of worked-out examples. To solve problems with the EA-Coach, students use the problem window (Fig. 1, left) to draw free body diagrams and type equations; students are free to enter steps in any order and/or skip steps. For each solution entry, the EA-Coach responds with immediate feedback for correctness, realized by coloring entries red or green, indicating correct vs. incorrect entries. Instead of providing hints, for instance on instructional material, the EA-Coach makes examples available to students (accessed with the "GetExample" button); these are displayed in the example window (Fig. 1, right). The system relies on a decision-theoretic approach to tailor the choice of example to a student's needs by considering problem/example similarity, a student's knowledge and reasoning capabilities (see [29] for details).

The study involved 15 participants, all Arizona State University students, who either were taking or had taken an introductory-level physics course. Each participant solved two physics problems with the EA-Coach of the type shown in Fig. 1; each problem solution involved about 15 steps (for further study details, see [25]). We used a variety of data collection techniques. First, the EA-Coach logged all interface actions. Second, we used talk aloud protocol [27]: we asked students to verbalize their thoughts; all sessions were taped and subsequently transcribed. Third, a sensor logger captured students' physiological responses from four sensing devices (see Fig. 2): (1) a *posture chair pad* measured position shifts (the pad included three pressure points on the chair seat and three on the back); (2) a *skin-conductance (SC) bracelet* captured skin conductance; (3) a *pressure mouse* measured the pressure exerted on the mouse (via six "pressure points"); (4) an *eye tracker* captured pupillary responses (the tracker was an integrated model that appeared as a regular computer screen).

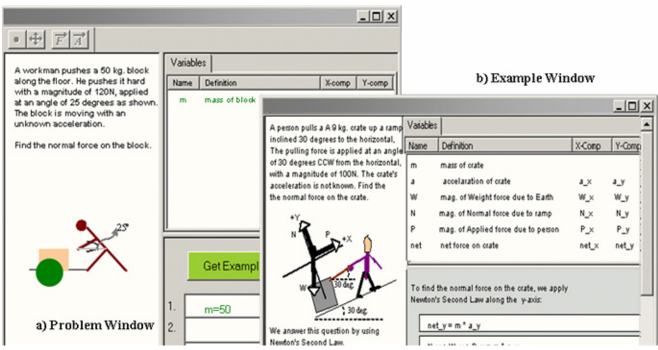

Fig. 1. EA-Coach problem and example windows

Fig. 2. Affective sensors (from left to right): posture chair, skin conductance (SC) bracelet, pressure mouse, Tobii eye tracker

3 Data Pre-processing

As our "gold standard" for "yes!" moments during the study, we relied on the verbal protocol data. Since a "yes moment" corresponds to excitement and/or positive affect, the transcripts were coded by the first author to identify such instances. As a starting point, we used data from an earlier affect coding [25], reanalyzing the codes to identify "yes!". We identified 68 "yes!" moments; all but one were directly associated with subjects generating a correct solution step and were expressed directly after doing so (recall that the EA-Coach provided immediate feedback for correctness so students were aware of the status of their entries). The one "yes!" that was not associated with a solution step occurred when a participant was reading the second problem statement after having already successfully solved the first problem.

While some of the "yes!" events were expressed in a very effusive manner ("*yes! I'm smart!", "oh yay*!"), others were more subdued ("*I got it right and that makes me feel good*"). In general, we found that when participants expressed a "yes!", it varied in terms of tone, expression, etc. Because we found it very difficult to disambiguate between the various forms of positive affect related to a "yes!", we decided to keep all instances in the analysis without trying to further distinguish between them.

We also had data on how subjects were reasoning during the study, obtained from an earlier coding [25] that included information on various types of reasoning, e.g., whether students were self-explaining by deriving physics principles, drawing comparisons between the problem/example, and/or expressing *some* form of cognitive processing (for examples, see [25]). For the purposes of this study, we collapsed the various types of reasoning into a single "*reasoning"* code, because as a starting point we were interested in how reasoning was related to "yes!" events.

Data Features. To analyze what events predict "yes!" moments, we identified a set of features we believed could be relevant. Note that the list presented here is not meant to be exhaustive, but rather to provide a starting point for understanding predictors of excitement/positive affect in instructional situations. First, we identified *interaction data* features we obtained from the EA-Coach logger corresponding to events in the tutor's interface, as follows:

- *Time*: The amount of time taken to generate a correct solution step (as described in Sect. 4, we focus on correct solution entries);
- *NumAttempts*: The number of attempts required to generate a correct solution step;

- *NumReasoning*: The number of "reasoning" utterances a student expressed in the process of generating a solution step;
- *Type of step*: The type of solution step (e.g., a force, an axis, an equation).

Second, we identified *sensor* features that we obtained from the sensor logger:

- *Pupillary response*: The mean change in pupil dilation around a point of interest (described in Sect. 4). For instance, if the point of interest is when a student generates a solution step, then mean change = (*mean pupil size over time span T directly following the step) - (mean pupil size over time span T directly preceding the step)*. We set the threshold *T=2* seconds, since this comparable to that used in other related work involving analysis of pupillary response (e.g., [30]).
- *Skin Conductance (SC)* response: The mean change in SC response around a point of interest (calculated as for pupillary response). We set the threshold *T*=2 seconds, based on the timeframe containing a SC response [14].
- *Mouse response*: The mean change in mouse pressure before and after an event of interest, using the method in [16] (where the mean pressure was obtained by summing over the pressure points, dividing by a constant and finding the mean). We set the threshold *T=10* seconds, because this sensor does not measure instantaneous responses (like SC and pupillary response) but rather longer scale transitions in behavior.
- *Chair*: The number of "*sitForward*" events, when a student leaned forward prior to generating a solution step, calculated by obtaining the pressure on the seat back via the formula in [16]. Here, we used a threshold *T=10* seconds, as for the mouse.

4 Results

In order to understand predictors of "yes!" in instructional activities, we compared "yes!" moments to other instances when students obtained a correct solution step but did not generate a "yes!". Since the "yes!" moments directly followed the generation of a correct solution step, we felt this would be the most appropriate comparison; this gave us 67 "yes!" instances[1] and 218 other events. As a final pre-processing step, for each logged correct step we extracted the above-described features, merging across the different log files (transcript, EA-Coach, sensor) to produce a single file.

Our hypotheses were that students would only express a "yes!" if they invested some effort into generating the solution step, and that there would be physiological manifestations of "yes!" that differed from other correct entries. To analyze whether these hypotheses were correct we carried out several types of analysis.

4.1 The Unique Nature of "yes!"

As a starting point, we wanted to determine if "yes!" moments differed from other correct entries (referred to as *other* below) in terms of the features listed above. Thus, we compared data on these two types of entries for our set of features through

[1] There was one exception where a student expressed "yes!" when reading an example; given our scheme, we did not consider this one data point in our analysis.

univariate ANOVA. As far as the *interaction* features are concerned, we found that students took significantly longer to generate a correct solution step corresponding to a "yes!" than other correct entries (on average, 206 sec. vs. 54 sec.; $F(1,297) = 77.27$; $p < 0.001$). Students also generated significantly more attempts for "yes!" entries, as compared to other correct entries (on average, 5.1 vs. 1.7; $F(1,297) = 40.47$, $p < 0.001$), and expressed significantly more reasoning episodes for "yes!" (on average, 1.34 vs. 0.58 $F(1,283) = 11.614$, $p = 0.001$). Our data was too sparse to analyze whether type of step had an effect.

As far as the sensor features are concerned, students had a significantly larger pupillary response for a correct solution step associated with a "yes!", as compared to other correct entries (on average, .043mm vs. -.037mm; $F(1,271)=8.422$, $p=0.004$). Skin conductance response had a marginal effect on "yes!" as compared to other entries (.000388μS vs. -.0000422μS, $F(1,291)=3.257$, $p=0.07$), suggesting a higher level of arousal for "yes!". Likewise, students had significantly fewer *sitForward* events before a "yes!", as compared to other entries (6.4 vs. 10.8; $F(1,296)=4.63$, $p=0.032$). One possibility for why this was the case is that students were more focused for "yes" entries and so were fidgeting less. We did not find "yes!" to have a significant effect on mouse response.

4.2 An Empirically-Based Model for Predicting "yes!"

The above analysis showed that "yes!" moments are uniquely distinguishable. To develop a user model, however, we need to understand how the various features predict "yes!" events. Thus, we conducted regression analysis. Because we have a nominal dependent variable ("yes!" vs. other), we used a logistic regression. A key consideration behind our choice of modeling technique was our data set size: while acceptable for modeling with logistic regression, where the rule of thumb is at least 20 data points per independent variable, it was not large enough for some other machine learning techniques, e.g., support vector machine. Of the applicable techniques, regression was chosen based on prior research showing its suitability for classifying affect ([16, 31]); [31] found that regression yielded the highest affect classification accuracy over other machine learning methods.

We begin by presenting the baseline model, one that always predicts the most likely event (here, lack of a "yes!"). Given the base rates of the two decision options and no other information, the best strategy is to predict that each step is not a "yes!". This model achieves 76% accuracy (# of correctly classified events / total # of events), but obviously completely misses our goal of predicting when "yes!" occurs (i.e., never predicts "yes!", and so has a true positive value of 0%, see Table 1, top).

Using the *Step* method, we then added our features to the logistic regression model[2]. The resulting model containing *time, numReasoning, pupil response, SC response* and *chair* was significantly better from the baseline model ($p<0.001$, see Table 1, top). Below, we will analyze the contribution of some of our features to the model's accuracy, but first we examine the full model accuracy.

---

[2] Because increasing the number of predictors decreases experimental power, we omitted *numTries* from this analysis, as it was redundant due to its high correlation with time; we also omitted *mouse response* since it did not significantly distinguish "yes!" from other entries.

Table 1. Logistic regression "yes!" models (TP=Sensitivity, TN=Specificity; Acc= TP + TN / *N*)

	Overall Logistic Regression Equation	**TP**	**TN**	**Acc.**
Baseline model	-2.206	0	100	76
Full model **	-2.206+*time**.008+ *numReasoning**.309 + *pupilResponse**1.68 +*SC**126.4+*chair**-.019	60.3	87.2	81.4
Time*+numReasoning*	-2.437 + *time**.01 + *numReasoning*.345*	55.2	89.2	81.6
Time*+pupilResponse*	-2.157+*time**.009+*pupilResponse**2.082	54.2	85.0	78.4
Time*+SC	-2.308 + *time**0.01+ *SC**128.97	57.6	89.9	82.6
Time*+Chair	-2.084 +*time**0.01 + *chair**-.017	56.7	88.3	81.2

*** Significantly better than baseline model, p<0.05*
** Each feature significantly improves model fit over previous model (i.e., model 1=baseline, model 2=time, model 3= time+2nd feature), p<0.05*

The output of a logistic regression equation is a probability that a given event belongs to a particular class. In order to use the model for prediction, it is therefore necessary to have a decision rule: if the probability of an event is greater or equal to some threshold then we will predict that event will take place (and not take place otherwise). To choose the optimal threshold, we built a Receiver Operating Characteristic (ROC) curve (Fig. 3). The ROC curve is a standard technique used in machine learning to evaluate the extent to which a classifier can successfully distinguish between data points (episodes correctly classified as positive, or *true positives*) and noise (episodes incorrectly classified as positive, or *false positives*), given a choice of different thresholds. Figure 3 shows the ROC curve we obtained for our "yes!" models, where each point on the curve represents a model with a different threshold value. As is standard practice, we chose as our final threshold the point on the curve that corresponds to a reasonable tradeoff between too many false positives vs. too few true positives (P=0.26, labeled by a cross on the curve in Fig. 3).

When reporting classifier accuracy, it is standard to provide *sensitivity* (true positives) and *specificity* (true negatives), since these are more informative then overall accuracy (true positives + true negatives / total number of instances). Our classifier is significantly better than the baseline model ($p < 0.05$) and obtained a sensitivity of 60.3%, a specificity of 87.2% (and overall accuracy of 81.4% - see Table 1, top). Thus, this classifier correctly identifies 60% of "yes" moments, without incorrectly classifying *other* entries as "yes!" for 87% of the time.

Model Validation. To validate the above model, we conducted a leave-one-out cross validation. Specifically, we trained the classifier using *N-1* data points and tested on the remaining data point, repeating this process *N* times (where *N* is equal to the number of samples, 269 full samples, i.e., without any missing data points that were the result of, for instance, the eye tracker failing to find a valid pupil reading). The validation showed that our model accuracy does not degrade substantially (i.e., sensitivity=55.2%, specificity = 87.1%, accuracy = 79.5%).

Parsimonious Models. We wanted to explore what kind of model fit we could obtain with a subset of our features, which helps to make an informed decision as to which sensors to use if not all are available. Thus, we ran a series of regressions using *time* as the tutor variable (as this variable was highly significant in our regression model)

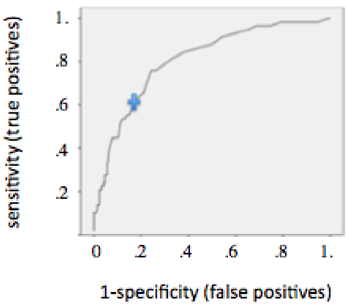

Fig. 3. ROC curve for various decision rule thresholds

and one of the other features. As Table 1 illustrates, we obtained reasonable results in terms of sensitivity and specificity with these reduced models, although only *reasoning* and *pupil response* resulted in significantly better models over a model that only included time (*SC response* and *chair* both improved the model fit, but this did not reach significance, i.e., p=.151 for the *SC response* and p=.153 for the *chair*).

5 Discussion and Future Work

In this paper, we reported on our analysis of moments of excitement and positive affect during instructional activities, which we refer to as "yes!" events. We found that "yes!" always followed a correct solution step, but conversely, a correct step was not always followed by a "yes!". In particular, students were significantly more likely to express a "yes!" after investing more time, generating more attempts, and expressing more reasoning episodes, as compared to correct entries for which corresponding enthusiasm was not expressed. Note that in addition to verbal expression of "yes!", another indication of arousal related to these events was provided by the pupil dilation and skin conductance data. These findings imply that students experience excitement and/or positive affect in tutoring situations when they have invested effort into the process and that effort pays off (i.e., correct solution is obtained). It is possible, however, that students express "yes!" not because they invested thoughtful, deliberate processing but because they guessed and/or arrived at the solution by luck. Our analysis does provide some indication that this is not the case, as students engaged in significantly more reasoning (captured by the "reasoning" code that included self-explanation, a form of deep processing) prior to "yes!". This does not guarantee every student behavior related a "yes!" is an instances of "deep" reasoning – in the future, we plan to delve deeper into this issue of mindful processing and "yes!".

To the best our knowledge, ours is the first work to propose a model for affect recognition incorporating pupillary response data. Although in contrast to the other low-cost sensors we used, eye tracking technology is more expensive, it is becoming more and more accessible, and so investigating its utility for user modeling is important. In a prior study [25], we also found a significant difference in pupil size between affective responses, but there are four key differences between that study and the present. First, in [25] we analyzed how pupillary response differs between positive and negative affect, without developing a model based on this data. Second, here we focus on "yes!" while in [25], we focused on differences between four affective states. Third, in [25] we normalized the pupil data using Z scores – while this approach is sometimes used (e.g., [19]) and increases experimental power, it requires subtracting the overall signal mean from each data point. Since this mean can only be obtained after a user finishes interacting with a system, the findings are difficult to apply for real-time user models. In contrast, here we use the raw signal values, making our findings more applicable to real-time modeling. Fourth, our feature set includes an array of sensors and tutor features, while in [25], we analyzed only pupillary data.

Overall, the tutor and sensor features resulted in a model that predicted "yes!" with 60% sensitivity and 87% specificity, a significant improvement over the baseline model. We also analyzed how using subsets of features impacts model fit: although the model incorporating the full set of features allowed the best trade-off between sensitivity and specificity, using a subset of features also resulted in models with reasonable fit. For instance, a model that includes only information on time and reasoning performs quite well – this may be useful if a system already has the tools to capture reasoning style (e.g., as in [32]) but sensors are not available. As far as the sensor features are concerned, when we explored parsimonious models, each sensor improved model fit over the time-only model. However, this improvement was only reliable, as reported by the p value, for the pupillary response feature. Compared to the pupil-based model, the models incorporating the other sensors resulted in higher specificity and/or specificity. These results, however, have to be interpreted with caution, since they approached but did not reach significance. This may be due to our modest sample size, and so more data is needed to confirm these sensors' utility.

While there is room for improvement, our model is a first step in providing information on "yes!" moments, which in turn can be used for tailoring pedagogical scaffolding to foster interest. For this purpose, it is key that the classifier not misclassify too many other entries as "yes!" (i.e., has high specificity), while still identifying *some* "yes!" moments, as is the case for our classifier. Given our limited sample size and particular instructional context, however, more work is needed to validate and generalize our findings.

Returning to our original four questions, we summarize the progress made so far and directions for future work.

1. *Are "yes!" events desirable outcomes or associated with desirable outcomes?* We argue that it is both. We now know that "yes!" occurs after students appear to have overcome a challenge related to generating a solution step, as indicated by time spent and number of tries produced. Since challenge fosters interest, this suggests that "yes!" events may be suitable as a predictor of increased learning, interest and motivation, something we plan to explore in future studies.
2. *What causes "yes!" events and how can we increase their frequency?* We now know that "yes!" events occur after a challenge is overcome with an example as the only the aid from the tutor. This is consistent with Lepper's advice of keeping the student optimally challenged [10].
3. *How can "yes!" events be useful to tutors, learning companions and other agents?* We offer some suggestions based on theory, but this remains to be empirically explored.
4. *Can "yes!" events be detected more accurately than a baseline approach?* Yes!

Acknowledgements. The authors thank the anonymous reviewers for their helpful suggestions and David Cooper, who contributed the sensor logging software. This research was funded the National Science Foundation, including the following grants: (1) IIS/HCC *Affective Learning Companions: Modeling and supporting emotion during learning* (#0705883); (2) *Deeper Modeling via Affective Meta-tutoring* (DRL-0910221) and (3) *Pittsburgh Science of Learning Center* (SBE-0836012).

References

1. Reeve, J.: The Interest-Enjoyment Distinction in Interest Motivation. Motivation and Emotion 13, 83–103 (1989)
2. Isen, A., Reeve, J.: The Influence of Positive Affect on Intrinsic and Extrinsic Motivation: Facilitating Enjoyment of Play, Responsible Work Behavior, and Self-Control. Motivation and Emotion 29(4), 297–325 (2005)
3. Hidi, S.: Interest and its Contribution as a Mental Resource for Learning. Rev. of Ed. Research 60(4), 549–571 (1990)
4. Isen, A., Daubman, K., Nowicki, G.: Positive Affect Facilitates Creative Problem Solving. J. of Personality and Social Psychology 52, 1122–1131 (1987)
5. Lepper, M.: Motivational Considerations in the Study of Instruction. Cognition and Instruction 5(4), 289–309 (1988)
6. Hidi, S., Renninger, A.: The Four-Phase Model of Interest Development. Educational Psychologist 41(2), 559–575 (2006)
7. Deci, E., Koestner, R., Ryan, R.: Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Again. Rev. of Ed. Research 71, 1–27 (2001)
8. Baker, R., D'Mello, S., Rodrigo, M., Graesser, A.: 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. J. of Human-Computer Studies (in press)
9. Boyer, K., Phillips, R., Wallis, M., Vouk, M., Lester, J.: Balancing Cognitive and Motivational Scaffolding in Tutorial Dialogue. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 239–249. Springer, Heidelberg (2008)
10. Lepper, M., Malone, T.: Intrinsic Motivation and Instructional Effectiveness in Computer-Based Education. In: Snow, R., Farr, M. (eds.) Aptitude, Learning and Instruction, vol. 3, pp. 255–296. Erlbaum, Hillsdale (1987)
11. Lepper, M., Woolverton, M., Mumme, D., Gurtner, J.: Motivational Techniques of Expert Human Tutors: Lessons for the Design of Computer-Based Tutors. In: Lajoie, S., Derry, S. (eds.) Computers as Cognitive Tools, pp. 75–105. Lawrence Erlbaum Associates, Hillisdale (1993)
12. Henderlong, J., Lepper, M.: The Effects of Praise on Children's Intrinsic Motivation: A Synthesis and Review. Psychological Bulletin 128(5), 774–795 (2002)
13. Chi, M., VanLehn, K., Litman, D., Jordan, P.: Inducing Effective Pedagogical Strategies Using Learning Context Features. In: Proc. of the 18th Int. Conference on User Modeling, Adaptation and Personalization (in press)
14. Kapoor, A., Burleson, W., Picard, R.: Automatic Prediction of Frustration. Int. J. of Human-Computer Studies 65(8), 724–736 (2007)
15. McQuiggan, S., Lester, J.: Diagnosing Self-Efficacy in Intelligent Tutoring Systems: An Empirical Study. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 565–574. Springer, Heidelberg (2006)

16. Cooper, D., Arroyo, I., Woolf, B., Muldner, K., Burleson, W., Christopherson, R.: Sensors Model Student Self Concept in the Classroom. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 30–41. Springer, Heidelberg (2009)
17. D'Mello, S., Graesser, A.: Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments. In: Luckin, R., Koedinger, K., Greer, J. (eds.) Proc. of the 13th Int. Conference on Artificial Intelligence in Education, pp. 161–168. IOS Press, Amsterdam (2007)
18. Conati, C., Zhou, X.: Modeling Students' Emotions from Cognitive Appraisal in Educational Games. In: Cerri, S.A., Gouardéres, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, pp. 944–954. Springer, Heidelberg (2002)
19. Conati, C., Maclaren, H.: Modeling User Affect from Causes and Effects. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 4–15. Springer, Heidelberg (2009)
20. Conati, C., Merten, C.: Eye-Tracking for User Modeling in Exploratory Learning Environments: An Empirical Evaluation. Know. Based Systems 20(6), 557–574 (2007)
21. Iqbal, S., Zheng, X., Bailey, B.: Task-Evoked Pupillary Response to Mental Workload in Human-Computer Interaction. In: Dykstra, E., Tscheligi, M. (eds.) Proc. of the ACM Conference on Human Factors in Computing Systems, pp. 1477–1480. ACM, NY (2004)
22. Marshall, S.: Identifying Cognitive State from Eye Metrics. Aviation, Space and Environmental Medicine 78, 165–175 (2007)
23. Vo, M., Jacobs, A., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., Hutzler, F.: The Coupling of Emotion and Cognition in the Eye: Introducing the Pupil Old/New Effect. Psychophysiology 45(1), 130–140 (2008)
24. Schultheis, H., Jameson, A.: Assessing Cognitive Load in Adaptive Hypermedia Systems: Physiological and Behavioral Methods. In: De Bra, P.M.E., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 225–234. Springer, Heidelberg (2004)
25. Muldner, K., Christopherson, R., Atkinson, R., Burleson, W.: Investigating the Utility of Eye-Tracking Information on Affect and Reasoning for User Modeling. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 138–149. Springer, Heidelberg (2009)
26. Conati, C., Maclaren, H.: Empirically Building and Evaluating a Probabilistic Model of User Affect. User Modeling and User-Adapted Interaction (in press)
27. Ericsson, K., Simon, H.: Verbal Reports as Data. Psych. Rev. 87(3), 215–250 (1980)
28. D'Mello, S., Craig, S., Sullins, J., Graesser, A.: Predicting Affective States Expressed through an Emote-Aloud Procedure from Autotutor's Mixed-Initiative Dialogue. Int. J. of Artificial Intelligence in Education 16(1), 3–28 (2006)
29. Muldner, K., Conati, C.: Evaluating a Decision-Theoretic Approach to Tailored Example Selection. In: Veloso, M. (ed.) Proc. of 20th Int. Joint Conference on Artificial Intelligence, pp. 483–489. AAAI Press, Menlo Park (2007)
30. Van Gerven, P., Paas, F., Van Merrienboer, J., Schmidt, H.: Memory Load and the Cognitive Pupillary Response in Aging. Psychophysiology 41(2), 167–174 (2001)
31. D'Mello, S., Picard, R., Graesser, A.: Towards an Affect Sensitive Auto Tutor. IEEE Intelligent Systems 22(4), 53–61 (2007)
32. Conati, C., VanLehn, K.: Toward Computer-Based Support of Meta-Cognitive Skills: A Computational Framework to Coach Self-Explanation. Int. J. of Artificial Intelligence in Education 11, 389–415 (2000)