Classifying Learner Engagement Through Integration of Multiple Data Sources

Carole R. Beal, Lei Qu, & Hyokyeong Lee
Information Sciences Institute, University of Southern California
4676 Admiralty Way, Marina del Rey CA 90292 U.S.A. < cbeal@isi.edu >
Keywords: Computer-aided education, Intelligent tutoring systems, Cognitive modeling

Abstract

Intelligent tutoring systems (ITS) can provide effective instruction, but learners do not always use such systems effectively. In the present study, high school students' action sequences with a mathematics ITS were machine-classified into five finite-state machines indicating guessing strategies, appropriate help use, and independent problem solving; over 90% of problem events were categorized. Students were grouped via cluster analyses based on self reports of motivation. Motivation grouping predicted ITS strategic approach better than prior math achievement (as rated by classroom teachers). Learners who reported being disengaged in math were most likely to exhibit appropriate help use while working with the ITS, relative to average and high motivation learners. The results indicate that learners can readily report their motivation state and that these data predict how learners interact with the ITS.

Learner motivation & tutoring systems

Technology-based instruction is becoming an important resource to improve learning outcomes in K-12 classrooms. Intelligent tutoring systems have been shown to improve learner achievement when used for supplemental instruction in the classroom (e.g., Koedinger, Corbett, Ritter, & Shapiro, 2000). Traditionally, tutoring systems have focused primarily on tracing students' knowledge states. For example, the Cognitive Tutor identifies and responds to misconceptions in students' solutions for algebra and geometry problems. However, there is growing recognition that student motivation and engagement must also be considered in addition to cognitive processes. Specifically, learners often do not use tutoring systems effectively. For example, in the case of mathematics tutoring systems, learners may choose random answers (guess), repeatedly request help until the correct answer is revealed (help abuse), or skip problems (avoidance).

Recent work has focused on the goal of attempting to estimate the probability that the learner is disengaged or is "gaming" the tutoring system by using time traces of student actions with the ITS (cf., Aleven & Koedinger, 2000; Beck, 2005). The results suggest that additional data sources may be useful in order to improve the ability of an ITS to diagnose the learner's goals: to use the system to learn, to solve the problems independently, or to game the system. More specifically, this approach has the potential to identify *when* students are disengaged, but does not help us to understand *why*, in terms of the beliefs and goals that individual students bring to the learning situation. Different students may appear disengaged for different reasons: One may act bored because he genuinely finds the work too easy; another may be capable of doing the work but lacks confidence and feels too anxious about failure to concentrate; and another may not have the required skills but is wary of using the ITS help because she has learned not to expect useful assistance from peers, parents or even teachers. It is unlikely that a single pedagogical response will be appropriate for all cases. Rather, the engagement tracing approach might well be enhanced by additional data sources about students' domain-specific expectations and learning goals.

The present research focuses on the integration of self-report data about learners' motivation with teacher reports of learner motivation and achievement, and classification of learner action patterns into finite-state machines. We decided to use student self-report data about motivation for two reasons: First, literally hundreds of studies in educational psychology indicate that learner motivation can be readily reported by learners, and that these data are strongly related to a cluster of behaviors associated with learner achievement, including effective self-monitoring, goal-setting, and study behaviors (for a review, cf., Schunk, 2004). In some studies, learner motivation and self-regulation are even stronger predictors of achievement than prior academic achievement or socioeconomic status (Byrnes, 2003; Zimmerman & Martinez-Pons, 1986). Thus, adding estimates of learner engagement could improve the effectiveness of tutoring systems, in terms of pedagogical decisions to restrict access to help or to force learners to view help.

Our second reason to evaluate the potential of student self-reports of motivation was more pragmatic. Much promising research focuses on the use of fragile, expensive and intrusive sensors (e.g., eye-tracking, skin conductivity, pressure sensors, etc.) to assess learner interest and engagement while using tutoring systems (cf., D'Mello et al., 2004). However, at present, this approach is limited to laboratory settings with small numbers of learners and cannot easily be used on wide scale in realistic educational delivery settings, such as the high school classes that use our mathematics ITS.

Another goal of the project was to evaluate the potential of classroom teachers as sources of expert knowledge about their students, in terms of both students' math ability, and mathematics motivation. Teachers have been primarily viewed as domain experts, meaning their knowledge of the content to be taught and the scaffolding to be provided. Yet teachers are able to assess students' performance, in terms of work on assignments, performance on tests, and apparent comprehension of the material; in fact, assessment is one of the primary functions of classroom teachers. Teachers are also able to judge students' interest in the material, their attitudes toward learning, their apparent effort to learn, and other behaviors indicative of motivation (Ryan, Patrick, & Shim, 2005). Thus, it seemed important to learn if teachers' assessments might predict students' productive or ineffective use of the ITS for learning.

Method and data sources

Tutoring system. The tutoring system was Wayang Outpost, a web-based application providing instruction in high school mathematics < www.wayangoutpost.net >. Students viewed a series of math word problems. Each problem included five answer options; students could choose an answer at any point and receive feedback (e.g., when an answer was selected, a red "X" or green checkmark indicated if the answer was right or wrong). Students could also request a multimedia explanation of the solution by clicking the "Help" icon. Explanations were constructed as an ordered sequence of individual hints leading to the correct answer. Individual hints included information presented in one modality (e.g., text, or animation, or audio) to avoid excessive cognitive load (cf., Mayer, Dow, & Mayer, 2003) but the complete explanation for a problem included hints with a range of modalities. Student actions were recorded in the server database, including clicks, sequence, and latencies between clicks.

Study participants. The study included high school students from three urban schools in Los Angeles. Students worked with the mathematics tutoring system as part of their regular classroom mathematics instruction. Data sets were available for 83 – 91 students, with some items missing for individual students due to absence on the day a particular task was administered.

Motivation Profile. In the first session, students completed an on-line self-report instrument used to assess mathematics motivation that was integrated into the tutoring system application. The on-line instrument was derived from integrating on-line and paper-and-pencil questionnaires previously shown to have high reliability and validity (Boekaerts, 2002; Eccles, Wigfield, Harold, & Blumenfeld, 1993). Because academic motivation is believed to be domain-specific, items were specific mathematics. The 10 item instrument included two questions addressing each of five constructs: Math self-efficacy; Beliefs that math is important to learn; Liking of math; Expected success in math; and Difficulty of math. The Motivation Profile also included an item designed to assess the learner's beliefs about math ability: "entity" beliefs reflect the view that math skill primarily reflects native ability, whereas "incremental beliefs" indicate the learner believes that skill can be enhanced through effort (Dweck, 2006). Students clicked on a Likert-type rating scale to indicate their answer. Answers were automatically recorded into the ITS server database.

Teacher ratings. The students' mathematics teachers provided categorical ratings of individual students' behaviors indicative of mathematics motivation and achievement in mathematics class. Motivation ratings included three categories: High self-regulation; grade-level (average) motivation; disengaged. Achievement ratings included three categories: Above grade-level; meeting grade-level expectations; below grade level, i.e., in danger of failing the class. All teachers were qualified mathematics teachers and each had more than 10 years of experience with high school math instruction.

Student data records. Data records were extracted from the ITS database for each student. A single student's data record consisted of a sequence of problem events, defined as the presentation of a problem, followed by the subsequent interface clicks (clicks on answers, requests for help) and latencies between clicks, terminated by the request for a new problem.

Action patterns. We defined five finite-state machines representing how students might work with the ITS. Rules for each are described below. Students' data records were machine scanned and each student's problem events were classified. (In the rules presented below, the limit of 10 seconds was generated after viewing a sample of traces of high-achieving students, on the grounds that if these skilled students required at least 10 seconds to read a problem or a hint, it was not likely that other students could do so in less time. Analyses conducted with 5 and 15 second windows yielded similar effects and interpretations.)

Independent-a. Problem is available for at least 10 seconds, followed by the selection of the correct answer.

We infer that the student read the problem and solved it correctly without ITS assistance.

Independent-b. Problem presented for at least 10 seconds, followed by an incorrect answer choice; another 10 or more seconds; followed by the correct answer. We infer that the student read the problem, computed an incorrect answer, and revised to the correct answer, without ITS assistance.

Guessing. Student selected one or more answers within 10 seconds of the problem presentation; no help was viewed. We infer that the student did not read the problem and clicked on answers until the correct answer is discovered.

Help abuse. The student clicked on "help" with inter-click intervals of less than 10 seconds. We infer that the student did not attend to the hint but was searching for the correct answer**.**

Learning. The problem was presented for at least 10 seconds; help was requested and presented for at least 10 seconds before an answer was selected or another hint was requested. We infer that the student read the problem and the help, i.e., was trying to learn how to solve the problem.

Results and Discussion

Learner motivation. Students' average scores for the five math motivation constructs (self efficacy, value, expected success, difficulty, and liking of math) were subjected to hierarchical cluster analysis yielding 3 groups. Mean scores for the three groups may be viewed in Table 1.

	Efficacy	Liking	Value	Diff.	Exp.Succ
Group1	3.19	2.47	3.36	2.75	3.37
Group2	1.89	1.60	3.14	2.58	1.90
Group3	3.91	3.86	4.79	3.83	4.16

Table 1: Mean scores on Motivation Profile by Group

Students in Group 1 (N = 50) appeared to have average motivation in math: they expected to pass, thought math might be important to their future, and liked it a bit less than other academic subjects. Group 2 students (N = 21) showed a distinctly different pattern: they did not like math and did not think they had much ability in math. Group 3 students (N = 12) had high beliefs in their ability, liked math much more than the other groups, and thought math was very important to learn.

Teacher ratings. Teacher ratings of motivation were highly correlated with students' self-reports of math motivation. A chi square analysis indicated that teacher motivation and achievement were significantly associated, e.g., students rated as performing above grade-level expectations also tended to be the same students who were rated by teachers as high in motivation. However, about 35% of the students were rated by teachers as having average to high motivation but also as being low in achievement, i.e., they were in danger of failing their math class.

Action patterns. Students completed an average of 31 math problems, with a range of 10 to 90. As may be seen in Figure 1, over 90% of students' problem events (totaling about 2,635) multi-step problems) could be classified into of the five action patterns. The remaining problems included cases of skipping or partial work on a problem (e.g., the student quit out of the tutoring application before completing a problem). Overall, students' behavior with the ITS could be described in terms of finite state machine representations.

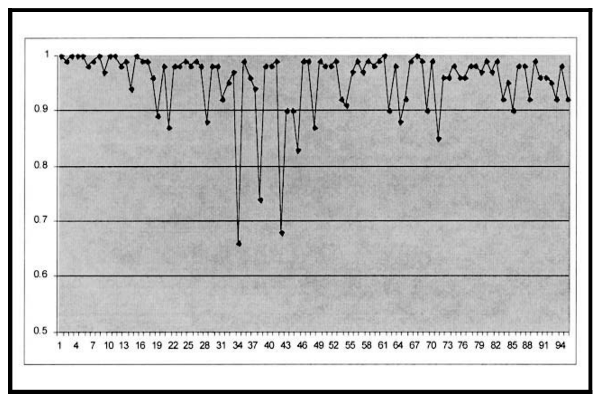

Figure 1: Summed proportion of math problems (Y axis) classified into Action Sequences, for individual students (listed on X axis)

The most common pattern was independent-a problem solving (32%); 21% were independent-b problems. Students were classified as attempting to learn on 22% of the problems. Guessing occurred on 19% of the problems, and help-abuse was observed on only 1% of the problems. More globally, students solved 53% of the problems on their own (Independent-a and –b combined); used help in an effort to learn on 22% of the problems; and "gamed" the system (Help-abuse and Guessing combined) on 20% of the problems.

Our work is thus consistent with others in that students' spontaneous use of the multimedia help available in the ITS was relatively low (Aleven & Koedinger, 2000). This is a concern because, first, students do not have the opportunity to benefit from the scaffolding if they do not ever access it and, second, from the knowledge engineering perspective, considerable resources go into the development of multimedia scaffolding, much of which students rarely see. However, to design a more effective pedagogical model, we need to have a deeper understanding of which students need to be encouraged to access help, which would benefit from being denied immediate answer feedback, and which students might learn best through other strategies, such as motivational as well as learning strategy feedback. Thus, our next analyses focused on the relation of student characteristics (motivational profile and achievement) with action patterns.

One question was whether we could proactively identify which students were most likely to solve the ITS problems independently, i.e., without incorrect answer attempts or accessing the multimedia help. Teachers' ratings of achievement were strongly predictive of students' independent problem solving, specifically, Pattern A, in which the student read the problem and chose the correct answer without error and without ITS assistance. A analysis of variance on students' Independent-a scores with teacher achievement (above-grade; grade-level; below grade-level expectations) as the grouping factor indicated that high achieving students solved significantly more problems on their own (47.8%) than average (39%) and low-achieving students (19%), $F(2,82) = 13.436$, $p < .001$. Although not particularly surprising, this result provides some indication that teachers' perception of their students' math knowledge was accurate, in that students who were independently rated as performing above grade expectations were more likely than other students to solve the ITS problems on their own. It also suggests that prior achievement could be used proactively to select a steeper difficulty curve for some students, rather than reactively in response to student problem solving behavior with the ITS, as reactive systems can lead to an overemphasis on easy problems (Arroyo, Murray, Beck, Woolf, & Beal, 2003). We are currently adding a feature into the ITS so that teachers can enter their ratings of student achievement before students begin to work with the system.

Teachers' assessments of their students' skills also matched students' own perceptions of how well they expected to do in math. An analysis of variance on students' scores for Expected Success (part of the Motivation Profile) with teacher achievement ratings as the grouping factor revealed that high-achieving students thought they would do better in math than lower-achieving students, $F(2,79) = 32.233$, $p < .001$. Mean ratings (from a range of 1 to 5) were 4.08 for high-achieving students, 3.31 for students with average achievement, and 2.47 for students who were performing below grade expectations. Thus, the end-users of the ITS (students in math classes) had a reasonably accurate sense of how well they were doing in math, and this self-assessment predicted an important aspect of their behavior with the ITS: independent and accurate math problem solving.

To investigate patterns in students' use of the ITS, proportion scores for the different patterns were used in a hierarchical cluster analyses yielding three groups. (Guessing and Help abuse scores were combined to produce one score, due to the low rate of Help Abuse.) Results may be viewed in Table 2.

	Guess	Learn	Ind.-a	Ind.-b
Group1	0.33	0.14	0.16	0.32
Group2	0.13	0.55	0.14	0.13
Group3	0.09	0.17	0.55	0.13

Table 2: Mean proportion for ITS action patterns by group

Group 1 students ($N = 33$) were most likely to guess while working with the tutoring system. Interestingly, these students were also likely to solve problems by making an incorrect guess and then viewing at least one hint (Independent-b pattern). Group 2 students ($N = 14$) were the highest users of the multimedia help for learning (55% of the problems). Group 3 students ($N = 36$) were most likely to solve the problems accurately and without using ITS help. These results indicate that individual students have systematically different strategies while using the tutoring system, and that these strategies can be described in terms of finite state machine components.

Our next question of interest was whether students' initial self-reports of motivation might explain which students subsequently adopted particular strategies as they worked with the ITS. A chi-square analysis on a cross-tab of students classified by motivation group and ITS action pattern group showed a significant relation, $^{2}(4,79) = 23.26$, $p < .001$.

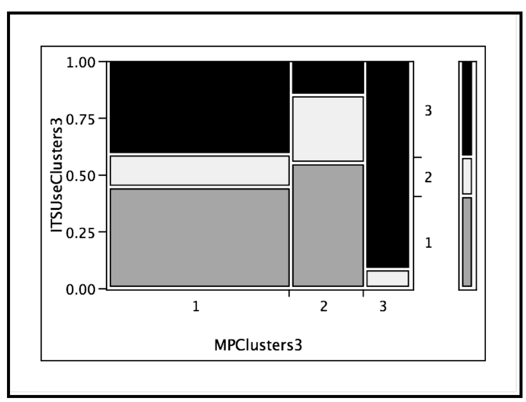

Figure 2: Mosaic plot of students by Math Profile Cluster and ITS Action Pattern Cluster

As may be viewed in Figure 2, students who had high math self-efficacy, liked math, and thought that math was very important to learn (Group 3) were most likely to solve the ITS math problems accurately and independently (right column). There were relatively few students who fit this description. In contrast, the largest group included students with average math motivation (left column) who seemed to be equally divided between those with a high proportion of guessing, and those who worked independently. More interesting was the group with low mathematics motivation (center column): About half had high guessing ratings, but the others had high learning action pattern rates. In fact, proportionally speaking, students who had low self efficacy, low attraction to math, and low expectations for success were *most* likely to use the ITS in a way that suggested an effort to learn, i.e., reading the problem and viewing the hints. The relatively high rate of learning-oriented ITS use by low motivation students suggests that technology-based instruction has potential to reach students who are not doing well with regular classroom instruction; such students are known to avoid seeking help from teachers and classmates (Karabenick, 1998; Newman, 2002; Turner et al., 2002). The opportunity to learn from software may offer an appealing alternative because the student can seek help in private.

Additional support for this interpretation was found in an analysis focusing on the average number of hints viewed per problem. An analysis of variance on mean number of hints per problem with Motivation Group as the grouping factor showed a significant effect, $F(2,82) = 20.525$, $p < .001$. Post hoc Tukey comparisons ($a = .05$) indicated that low motivation students viewed an average of 1.94 hints per problem, whereas average and high motivation students viewed fewer hints (0.37 and 0.34, respectively). This result suggests that the low motivation students were not only more likely to use the ITS resources, they drilled deeper into the resources once accessed by viewing multiple hints. Of course, the high motivation students were also more likely to successfully solve problems independently, indicating that they did not need to view the help to succeed. Again, however, the point is that the ITS may offer low motivation students an effective way to improve their problem solving skills, and they seem willing to accept this opportunity. This is especially striking given that the ITS allowed them to progress through the problems by guessing, yet it was the average motivation students who were more likely to follow this strategy.

Having established that students' motivation helped to predict patterns of actions with the ITS, a related question was whether we could evaluate the relative contribution of students' motivation and their prior math achievement (based on teacher ratings). A logistic regression was conducted on the cross-tab of student motivation group and teacher rating group (high, grade-level, or below-grade-level achievement), with action pattern strategy group as the outcome factor. The results for the whole model indicated a significant lack of independence, $^2(8) = 30.945$, $p < .001$. This means that the relative proportion of guessing, learning and independent problem solving exhibited by individual students as they worked with the ITS was not independent of their motivation or achievement in math (as rated by their teachers). More significantly, likelihood ratio effects tests indicated that motivation group membership contributed to the model, $^2(4) = 12.055$, $p < .01$, but that teacher achievement rating did not, $^2(4) = 7.737$, $p = 0.12$, N.S. This supports our claim that student motivation – the constellation of beliefs about one's ability and the value of learning the subject – must be considered in the design of tutoring systems, in addition to the more traditional focus on cognitive modeling.

We attempted to learn more about the largest group of students: those with average mathematics motivation ($N = 50$) who were roughly split between those who tended to guess (44%) and those who tended to solve ITS problems independently (40%). One might expect that, within this subgroup, the independent problem solvers would be those with higher math skills, yet teacher achievement ratings did not predict ITS strategy for this group of students. However, the learning orientation item on the Mathematics Profile was predictive to some extent. Specifically, students with average mathematics motivation who had higher guessing rates believed that math ability is fixed, whereas their peers who worked independently believed that math ability can be increased through effort, $^2(2) = 8.812$, $p < .05$. Our interpretation of this result is limited because only one item was used to assess students' beliefs about the role of native ability versus effort in learning mathematics. Still, this finding is consistent with our view that students' self-reported beliefs about learning predict aspects of their behavior while using an ITS, independent of their prior achievement in the domain.

Conclusions

We have shown that multiple data sources can be integrated and used to classify students in terms of the constellation of beliefs that they bring to the learning situation. These data were readily acquired from users, and were consistent with teachers' knowledge of their students' achievement and motivation. The classifications also predicted students' strategies while using the ITS, particularly their tendency to guess, to work independently, or to use the multimedia help to learn. In addition, our empirical approach led to the identification of students who described themselves as disengaged and discouraged about their ability to learn math, but who were at least as likely (and in some cases more so) as other students to use the

ITS in a manner suggesting they were attempting to learn. The results indicate that it should be possible to seed pedagogical models in advance with learner profile data that is timely and inexpensive to elicit, and quite predictive of strategies that will be employed once students begin working with the ITS.

Future work

The next step in our project is to link these data sources with a model of pedagogical feedback that is based on studies of expert human tutors (Lepper, Woolverton, Mumme, & Gurtner, 1993). Experienced human tutors continually balance the goal of helping the student learn new material with the goal of maintaining the student's motivation to learn, and accomplish this balancing act through a repertoire of feedback messages, sophisticated problem selection, and judicious offers of learner control when the learner appears to be flagging. Our ITS pedagogical model uses the action patterns identified in this study to classify the current student's most likely strategy and then selects suitable messages (e.g., encouraging the student to increase effort, or pointing out that the use of help in a previous problem led to success on the current problem) from a bank of messages organized in relation to motivational profile and achievement. An experimental evaluation is being conducted to compare the strategies and learning outcomes of students who work with the ITS with the enhanced pedagogical model, or a traditional version of the same ITS.

References

Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), Proceedings of the 5th International Conference on Intelligent Tutoring Systems, pp. 292-303. Amsterdam: IOS Press.

Arroyo, I., Murray, T., Beck, J. E., Woolf, B. P., & Beal, C. R. (2003). A formative evaluation of AnimalWatch. Proceedings of the 11th International Conference on Artificial Intelligence in Education, pp. 371-373. Amsterdam: IOS Press.

Beck, J. (2005). Engagement tracing: Using response times to model student disengagement. In C. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology, pp. 88-95. Amsterdam: IOS Press.

Boekaerts, M. (2002). The on-line motivation questionnaire: A self-report instrument to assess students' context sensitivity. In P. R. Pintrich & M. L. Maehr (Eds.), Advances in Motivation and Achievement, Vol. 12: New Directions in Measures and Methods, pp. 77-120. New York, JAI (Elsevier Science).

Byrnes, J. P. (2003). Factors predictive of mathematics achievement in White, Black and Hispanic 12th graders. Journal of Educational Psychology, 95, 316-326.

D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R., & Graesser, A. C. (2004). Integrating affect sensors into an intelligent tutoring system. In Affective Interactions: The Computer in the Affective Loop. Proceedings of the 2005 International Conference on Intelligent User Interfaces, pp. 7-13. New York: AMC Press.

Dweck, C. (2006). Mindset. New York: Random House.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self and task perceptions during elementary school. Child Development, 64, 830-847.

Karabenick, S. (1998). Strategic help seeking: Implications for learning and teaching. Mahwah NJ: Erlbaum.

Koedinger, K. R., Corbett, A. T., Ritter, S., & Shapiro, L. J. (2000). Carnegie Learning's Cognitive Tutor: Summary of research results. White Paper. Pittsburgh PA: Carnegie Learning.

Lepper, M. R., Woolverton, M., Mumme, D., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), Computers as cognitive tools, pp. 75-105. Hillsdale NJ: Erlbaum.

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining Environment: What works in the design of agent-based microworlds? Journal of Educational Psychology, 95, 806-812.

Newman, R. S. (2002). How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. In S. Pape, B. Zimmerman, B., & F. Pajares (Eds.), Theory into practice: Special issue: Becoming a self-regulated learner, pp. 132-138. Columbus, OH: The Ohio State University.

Ryan, A. M., Patrick, H., Shim, S-O. (2005). Differential profiles of students identified by their teacher as having avoidant, appropriate, or dependent help-seeking tendencies in the classroom. Journal of Educational Psychology, 97, 275-285.

Schunk, D. H. (2004). Learning theories: An educational perspective (4th ed.). Upper Saddle River NJ: Pearson.

Turner, J. C., Midgley, C., Meyer, D. K., Dheen, K., Anderman, E. M., Kang, Y, & Patrick, H. (2002). The classroom environment and students' reports of avoidance strategies in mathematics: A multimethod study. Journal of Educational Psychology, 94, 88-106.

Zimmerman, B., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self regulated learning strategies. American Educational Research Journal, 223, 614-628.

Acknowledgments

The Wayang Outpost project has been supported by National Science Foundation grants HRD 0120809, HRD 0411532, and REC 0411886. The views expressed are not necessarily those of the funding agency. We would like to thank the principals and teachers in the Pasadena Unified School District and the Bravo Medical Magnet High School in the Los Angeles Unified School District for their support of the research. We would also like to thank project members Erin Shaw, Hannes Vilhjalmsson, and Eusebius Lai.