

# A Multimedia Adaptive Tutoring System for Mathematics that Addresses Cognition, Metacognition and Affect

Ivon Arroyo · Beverly Park Woolf ·  
Winslow Burelson · Kasia Muldner · Dovan Rai ·  
Minghui Tai

Published online: 30 September 2014

© International Artificial Intelligence in Education Society 2014

**Abstract** This article describes research results based on multiple years of experimentation and real-world experience with an adaptive tutoring system named Wayang Outpost. The system represents a novel adaptive learning technology that has shown successful outcomes with thousands of students, and provided teachers with valuable information about students' mathematics performance. We define progress in three areas: improved student *cognition*, *engagement*, and *affect*, and we attribute this improvement to specific components and interventions that are inherently *affective*, *cognitive*, and *metacognitive* in nature. For instance, improved student *cognitive outcomes* have been measured with pre-post tests and state standardized tests, and achieved due to personalization of content and math fluency training. Improved student *engagement* was achieved by supporting students' metacognition and motivation via affective learning companions and progress reports, measured via records of student gaming of the system. Student *affect* within the tutor was measured through sensors and

---

I. Arroyo (✉)

Learning Sciences and Technologies, Department of Social Sciences and Policy Studies, Worcester Polytechnic Institute, Worcester, MA, USA  
e-mail: iarroyo@wpi.edu

B. P. Woolf

School of Computer Science, University of Massachusetts, Amherst, MA, USA

W. Burelson

New York University College of Nursing, New York University, New York, NY, USA

K. Muldner

Institute of Cognitive Science, Carleton University, Ottawa, Canada

D. Rai

Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA, USA

M. Tai

School of Education, University of Massachusetts, Amherst, MA, USA

student self-reports, and supported through affective learning companions and progress reports. Collectively, these studies elucidate a suite of effective strategies to support advanced personalized learning via an intelligent adaptive tutor that can be tailored to the individual needs, emotions, cognitive states, and metacognitive skills of learners.

**Keywords** Intelligent tutoring · Cognition · Metacognition · Affect · Mathematics education

## Introduction

Intelligent tutoring systems, Cognitive Tutors, and adaptive learning environments are all variations of the same common theme: instructional systems that contain empirical models of the student to *predict* student behaviors and knowledge, and to *act* upon these predictions to make pedagogical moves as students progress towards gaining expertise and mastery of the target domain. Such systems have expanded across many critical and complex domains with far reaching results (Woolf et al. 2009). These systems have been embedded within formal education, in K-12 schools (e.g., Arroyo et al. 2009a; Gobert et al. 2013; Heffernan et al. 2012; Koedinger et al. 1997) as well as in military/corporate training and in organizational learning (Lajoie and Lesgold 1989; Stottler and Domeshek 2005). They offer learners various affordances, such as personalized trajectories of content, just in time feedback, and flexibility in terms of progressing through the instructional materials.

These systems have been created for a variety of domain areas. Within the specific area of mathematics education, the challenges to be addressed by an adaptive tutoring system in order to be successful within the school system are numerous. Such systems have to support students as they move throughout the K-12 school system, as they develop expertise with numbers and operations, measurement and data, statistics and probability, geometrical concepts in 2D and 3D, in algebra and equations, relationships and functions. Traditionally, teachers model new concepts in front of the class, engage students in a variety of activities, and accompany, support, and help students through their development of abstract thinking.

However, math teachers face many challenges related to supporting the varied skill and motivation levels in their classrooms, where some students excel while others lack abilities for their corresponding grade level (e.g., in literacy or numeracy). Many students reach higher-level classes missing the foundations of mathematical reasoning such as rational numbers (e.g., fractions) or even basic whole number arithmetic. The math difficulties of individual students vary across students within a class, making it very challenging for mathematics teachers to meet the needs of every student. Beyond weaknesses in specific knowledge components, students often face common challenges, such as difficulty with number sense, i.e., a sense of quantity, a sense of magnitude and length, and the ability to flexibly manipulate numbers through various operations (Mazzocco et al. 2011); difficulty in transferring mathematical knowledge to novel contexts, difficulty addressing even small variations within the same context (Carraher et al. 1985; Carraher and Schliemann 2002); and difficulty making reasonable estimations (Lemaire and Lecacheur 2011). Some students with low achievement in mathematics struggle with memory tasks, such as quick retrieval from long term memory, which affects their mathematical fluency (Tronsky and Royer 2003), as well

as working memory capacity and executive function (Bull and Scerif 2001). These are very basic cognitive abilities that improve as children develop and are key to solving any kind of non-trivial mathematical task (Geary and Hoard 2003; Royer 2003), but low achieving students, including ones with math disabilities, generally have difficulty in these basic skills and need extra support and practice. These students often become disengaged, and fail to develop and implement the metacognitive skills that would be most effective to their learning, eventually developing a more negative affective relationship to mathematics than their peers, and reporting more negative emotions towards math problem solving (Woolf et al. 2010). Still, most of the literature in mathematics learning difficulties has focused on understanding where and how students struggle but research has rarely deeply analyzed how to address these difficulties, at least not in an extensive way. Thus, understanding effective means of supporting low achieving students and those who struggle with mathematics is a key goal of the Wayang research agenda, and fills an important gap in this area of research.

Students' individual mathematical differences, their strengths and weaknesses in specific areas of the curriculum or basic cognition are accompanied by various other challenges. These relate to, for instance, the regulation of students' own learning, the when, how and why to carry out a variety of self-regulatory actions in the course of learning and practicing. Some self-regulatory actions require accurate personal judgment of knowledge and learning, accompanied by important decisions related to effective help seeking (Aleven et al. 2003), understanding which problem solving strategies to apply, and setting manageable goals and objectives for one-self (Locke and Latham 2002). Research has repeatedly shown that such *metacognitive* and *self-regulative behaviors* are major factors influencing students' academic success. In a meta-analysis of self-regulatory training programs, (Dignath et al. 2008) summarized evidence that by addressing self-regulation during learning, students' academic performance, strategy use, and motivational outcomes were improved. The analyzed scaffolds were all human-based, and only few attempts are now beginning to appear for computer-based metacognitive support (e.g., Azevedo et al., 2011; Long and Aleven 2013). Even that work is specifically dedicated to studying self-regulation interventions on student learning, but more research is needed to understand how educational technologies that tailor metacognitive support to individual student needs, can weave the complex interleaving of self-regulation, learning and affective outcomes, at each step of the learning process. More research is needed to effectively customize metacognitive support for individuals of various levels and abilities to tailor tutor responses to individuals and groups that seem to respond differently to such support. The Wayang Outpost research is seeking to address this gap.

Another major factor that influences students' learning is their general *affective experience* as they learn. For instance, experiences of confidence, boredom, and confusion are major aspects influencing students' academic success and predictors of achievement (Pekrun et al. 2007, 2010). Certain affective experiences can hinder learning by increasing unproductive behaviors (e.g., Baker et al. 2004). While students' affective experiences may be both a cause and a consequence of specific non-productive actions towards learning and achievement, nonetheless affect plays a critical role in education, both in short term performance outcomes and in long term life-long career choices. Note that in many ways, affective experiences such as interest and curiosity towards math and science are more important for long-term outcomes than are the short-term outcome of mastering a specific knowledge unit (e.g., see Ceci et al.

2009; Royer and Walles 2007). Research in computer-based interventions so far has focused on short-term measures of success (e.g., correct answers) and longer-term measures (e.g., interest in mathematics and value of mathematics) have been neglected. Long-term measures have always been part of the Wayang research agenda. The next generation of intelligent learning systems should examine longer-term outcomes of affect, as well as examine how to juggle the complexities of dealing with competing outcome measures, while optimizing both student affect and cognitive mastery.

The learning sciences community acknowledges the importance of affect in students' experiences with educational systems and has developed technologies that can automatically detect and respond to student affect (Arroyo et al. 2009a; Calvo and D'Mello 2010; Conati and Maclaren 2009; Cooper et al. 2010; Graesser et al. 2007; McQuiggan et al. 2008; Muldner et al. 2010). As mentioned above, intelligent tutoring systems have to carry out two major tasks: *model* the student and *act* upon estimates of student states. Modeling affect is a critical first step for providing adaptive support tailored to students' affective needs; however, beyond modeling, to date little work exists that systematically explores the impact of using model information to tailor affective interventions and explores the impact of doing so on students' performance, learning, motivation and attitudes (i.e., how to respond to students' emotions, such as frustration, anxiety, boredom and hopelessness), as well as how to help students to regulate their learning process, within one comprehensive learning environment.

The research described in this article aims to address all of the areas discussed above, acknowledging the complexity of the mathematics education problem, where each individual has strengths and weaknesses in various areas. In many ways, we follow the ideas of Allen Newell, who in this excerpt expresses how intelligent tutoring systems are a step in the direction of smart machines for education:

Exactly what the computer provides is the ability not to be rigid and unthinking but, rather, to behave conditionally. That is what it means to apply knowledge to action: It means to let the action taken reflect knowledge of the situation, to be sometimes this way, sometimes that, as appropriate.... In sum, technology can be controlled especially if it is saturated with intelligence to watch over how it goes, to keep accounts, to prevent errors, and to provide wisdom to each decision.

— Allen Newell, *Fairy Tales*, AI Magazine, Vol 13. Number 4. 1992.

In particular, this article focuses on one landmark mathematics adaptive tutoring system that takes a student-centered and holistic approach to teaching, by addressing *cognitive*, *metacognitive* and *affective* factors that influence students as they learn. This system is the Wayang Outpost Mathematics Tutoring System,<sup>1</sup> now referred to as MathSpring, an intelligent tutor for mathematics that uses a variety of computational techniques to promote student learning in meaningful, effective, and efficient ways.

The key open questions that we seek to address are: 1) How to design tailored interventions within educational technologies that encourage students to feel positive about their learning experience? (see “[Affective Scaffolds and Interventions](#)” section: Affective scaffolds); 2) How to tailor tutor behavior to support students to self regulate their learning? (see “[Metacognitive Scaffolds and Interventions](#)” section:

<sup>1</sup> Wayang Outpost is now called “MathSpring” and is built in HTML, see [MathSpring.org](http://MathSpring.org).

Metacognitive scaffolds); 3) How to support personalized trajectories through the mathematics content, adjusting difficulty and fostering fluency over pre-requisite knowledge components (see “[Cognitive Scaffolds and Interventions](#)” section: Cognitive scaffolds); 4) How to build interventions for one construct that also provides cross-over to other constructs (e.g., cognitive interventions impact affective outcome; affective interventions impact metacognitive outcome); and 5) Can a single tutoring system effectively address all three constructs?

Few if any other systems address aspects of all three constructs: affect, metacognition, and cognition simultaneously, so the integration and efficacy of attending to these collectively is likewise an open issue. The challenges related to addressing these questions pertain to the fact that it is unclear how and when to tailor scaffolding of student affect in educational technologies, or when to stop to think about ‘why’ and ‘how’ students are using the software to provide metacognitive support, or when to provide further cognitive challenges and increased difficulty of math activities instead of going back to review pre-requisite knowledge components.

Our approach in addressing all these challenges has been to employ a student-centered design process by iteratively piloting interventions, refining them and performing large scale evaluations, both of individual components and of the system as a whole, in order to obtain a holistic view of their pedagogical utility over a variety of cognitive, affective and metacognitive outcomes.

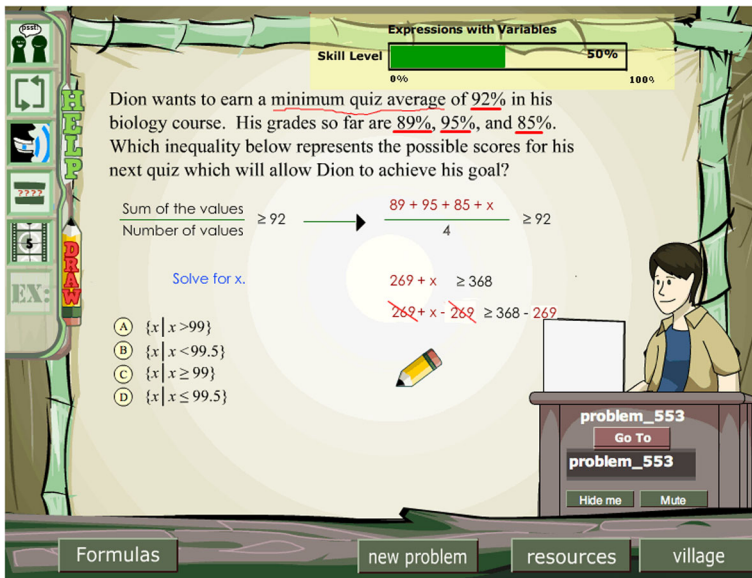
### Wayang Outpost

Wayang Outpost is a multimedia-based intelligent tutoring system (Woolf 2009) that provides a broad range of pedagogical support while students solve mathematics problems of the type that commonly appear on standardized tests, (see Fig. 1 and 2) for examples (Arroyo et al. 2004).

The tutor supports strategic and problem-solving abilities based on the theory of cognitive apprenticeship (Collins et al. 1989) that take place when a master teaches skills to an apprentice. In this case, the expert is the Wayang system that assists students during mathematics problem solving. The software models solutions via worked-out examples with the use of sound and animation, and provides practice opportunities on math word problems. Math problems are mostly released items from the Massachusetts Comprehensive Assessment System (MCAS),<sup>2</sup> administered annually to K-12 students as standardized tests; items are also generated by teachers and researchers, and from the Scholastic Achievement Tests (SAT)<sup>3</sup> - Math and other state-wide standardized tests across the country. Our choice of focusing on items of such high-stakes tests are mainly two-fold: 1) the items are good and non-trivial, as they require deep processing and strategic thought, beyond a rote procedure; 2) passing these tests has huge implications on a

<sup>2</sup> MCAS exams are given to all public school students in Massachusetts, on English Language Arts, Mathematics, and Science and Technology/Engineering. Students must pass the grade 10 tests in English Language Arts (ELA), Mathematics and one of the four high school Science and Technology Engineering tests as one condition of eligibility for a high school diploma.

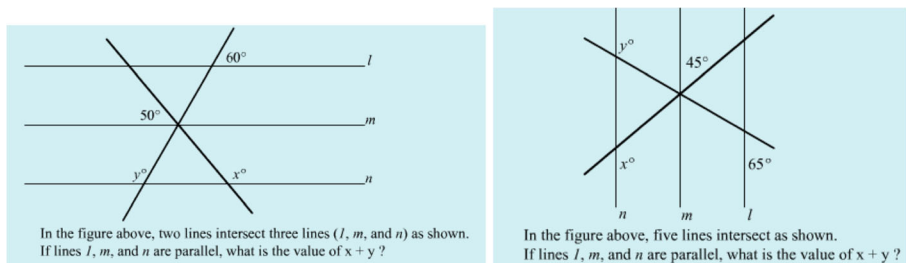
<sup>3</sup> The SAT is a globally recognized college admission test designed to assess a student’s academic readiness for college. These exams are typically given to graduating high school students and the results are used for financial support and scholarships. Many colleges use SAT Tests for admission, course placement, and to advise about course selection.



**Fig. 1** The Wayang Outpost Math Tutor interface. An animated companion provides individualized comments and support

student's future, thus they are high-stakes, for example students who don't pass the MCAS grade 10 test by the end of high school cannot graduate from high school in the United States, and students who don't score reasonably high in the scholastic aptitude test will not be accepted into good colleges. Many classes are given regularly to support students and remedial students to re-take these tests (e.g., the students from year 2005 in Fig. 4 were preparing to take their 10th grade MCAS test for a second or third time).

Using high-stakes test items as part of the teaching content of the tutoring system has implications for the design of the system, and especially on how the knowledge components are structured. Compared to most other math tutoring systems, the granularity of knowledge units/components in Wayang Outpost is more coarse-grained, and the differences across items within a knowledge unit is larger, with only pairs or triplets of problems being highly similar and parameterized (e.g., changing only operands, or words in the text). Instead, problems from high-stakes tests (released items from MCAS and SAT) are clustered together during a content-organization phase, generating sets



**Fig. 2** Two math problem items that involve exactly the same mathematical procedure to solve, but have a different difficulty level — the one on the right is estimated to be more difficult than the one on the left



that involve similar math skills. This is an important difference from other mathematics tutoring systems that may benefit students, as the implicit skill that is being addressed is an identification and classification task – students must identify what *kind* of problem this is, and how to approach it, within a certain set of problems that involve a subset of math skills. This classification/identification task is a major skill involving analogical mapping that can help students succeed at standardized high-stakes tests, where math problems can “look” very different from each other. We argue that the kinds of mathematical and abstract problems that students must solve in real life are varied too, thus, this classification/identification phase seems a critical part of mathematics education.

For instance, Wayang has a knowledge unit of called *decimals and percentages*. A student will know they are working on *decimals and percentages*, still the student needs to identify whether the problem they are solving involves converting from a decimal to a percent, or from a fraction to a decimal, or from a fraction to a percent, or requires interpreting a percent, or computing the percent of a number. We find that, if students see a very similar problem again that is simply parameterized with other operands/words, the student automatically identifies the problem making it much easier to solve –eliminating an important problem type classification task that is essential in high-stakes tests and in application of math in real life situations in general.

Scaffolds or hints are a key component of the Wayang Outpost tutor to help students learn strategies to approach math problems; hints are spoken to showing students steps towards a solution and implementing principles from multimedia learning theory (Mayer 2001), such as the contiguity principle, modality principle, animation principle, etc. Wayang Outpost is particularly strong at coaching and scaffolding, as it provides synchronized sound, animations, contiguous explanations using the math problem space (e.g., underlining, drawing on the figure), videos that show instructors solving problems and graphically provides virtual pencils to support student problem solving with their own notes, if so desired. It also provides worked-out examples as scaffolding (gives a worked-out example of a problem similar to that presented on the screen).

An important element of cognitive apprenticeship is to challenge students by providing slightly more difficult problems than learners/apprentices could accomplish by themselves. Vygotsky (1978) referred to this as the zone of proximal development and suggested that fostering development within this zone led to the most rapid learning (Murray and Arroyo 2002). The software provides adaptive selection of problems with increased/decreased difficulty depending on recent student success and effort (Arroyo et al. 2010a; Corbett and Anderson 1995). While there is no explicit “scaffold fading” procedure in Wayang Outpost, ‘help fading’ happens naturally as students first learn from the help provided in one math problem, and then are given a new problem of similar difficulty, to encourage them to transfer and perform problems without the need of help. This “help fading” process therefore happens naturally as the system encourages transfer of student knowledge to subsequent questions of similar difficulty (rows 2, 4, 9 within Table 1), until the student demonstrates mastery and challenge is increased. Later sections will show that Wayang also includes motivational learning companions that act much like students’ peers, offering affective support during the learning process and supportive talk as students become disengaged, see Figs. 14 and 15.

**Table 1** The effort-based tutoring algorithm informs pedagogical moves and affective decisions (last two columns) for each student on each problem. The algorithm first infers a reason for students behavior (fourth column) based on the number of incorrect student answers, hints requested and the amount of time spent (first three columns). Then the algorithm decides which pedagogical action the tutor should take (last two columns). The algorithm encourages transfer of student knowledge to subsequent questions of similar difficulty (rows 2, 4, 9), encouraging students to transfer skills and “fade” their need for help

Observed behavior and inferred reason for this behavior				Pedagogical Model Moves Cognitive or Affective or Metacognitive		
Incorrect	Hints	Time	Most Likely Reason	Decision	Moves	Cognitive or Affective or Metacognitive
1 $< E(Ti) - \delta_{iL}$	$< E(Hi) - \delta_{iL}$	$< E(Ti) - \delta_{iL}$	Mastery without effort	Increase Problem Difficulty	Show learning progress	Decisions
2 $< E(Ti) - \delta_{iL}$	$< E(Hi) - \delta_{iL}$	$> E(Ti) + \delta_{iH}$	Mastery with high effort	Maintain Problem Difficulty	Affective feedback: Praise Effort	
3 $< E(Ti) - \delta_{iL}$	$> E(Hi) + \delta_{iH}$	$< E(Ti) - \delta_{iL}$	Hint abuse, low effort	Reduce Problem Difficulty	Deemphasize importance of immediate success	
4 $< E(Ti) - \delta_{iL}$	$> E(Hi) + \delta_{iH}$	$> E(Ti) + \delta_{iH}$	Towards mastery, effort	Maintain Problem Difficulty	Praise effort	
5 $> E(Ti) + \delta_{iH}$	$< E(Hi) - \delta_{iL}$	$< E(Ti) - \delta_{iL}$	Quick guessing, low effort	Reduce Problem Difficulty	Deemphasize importance of immediate success	
6 $> E(Ti) + \delta_{iH}$	$< E(Hi) - \delta_{iL}$	$> E(Ti) + \delta_{iH}$	Hint avoidance and high effort	Reduce Problem Difficulty	Offer hints upon incorrect answer in the next problem	
7 $> E(Ti) + \delta_{iH}$	$> E(Hi) + \delta_{iH}$	$< E(Ti) - \delta_{iL}$	Quick guess and hint abuse	Reduce Problem Difficulty	Deemphasize importance of immediate success	
8 $> E(Ti) + \delta_{iH}$	$> E(Hi) + \delta_{iH}$	$> E(Ti) + \delta_{iH}$	Low mastery and High Effort	Reduce Problem Difficulty	Emphasize importance of effort and perseverance	
9 Otherwise	Expected Behavior	Maintain Problem Difficulty				



## Teacher Tools

In addition to supporting students, Wayang Outpost provides support for teachers. In particular, Wayang's assessment of student performance throughout their interaction with the system is reified through a graphical interface designed to show student progress to teachers.

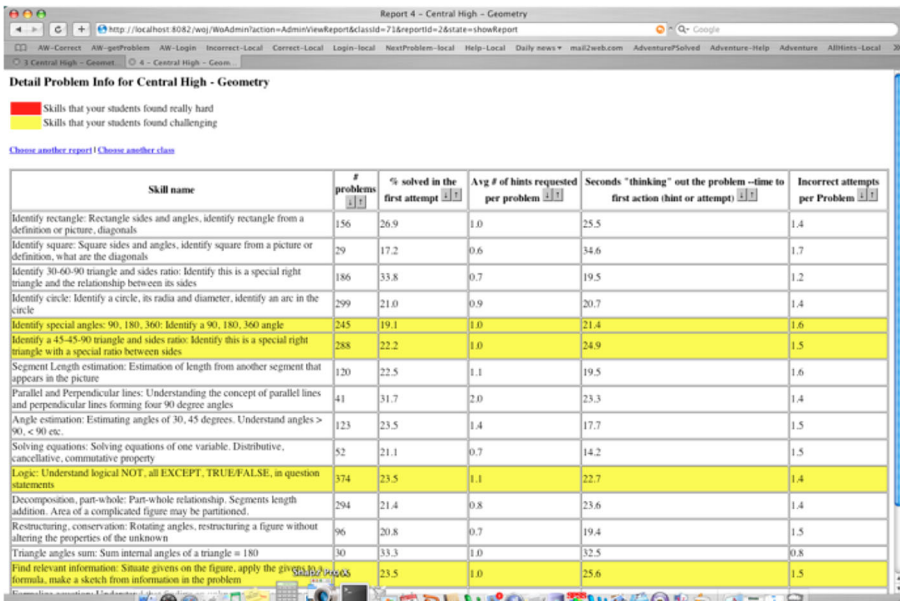
The teacher's interface provides several reports, including estimates of student learning overall, as well as progress per individual knowledge unit, assessments over individual problems as well as over math skills, and assessments aggregated by class, or by student. Wayang micro-analyzes features related to deep learning, such as fine-grained behavior while students solve problems related to timing (e.g., time to attempt an answer or read a problem, amount of help requested, etc.). This data is delivered to teachers immediately in real-time as students work on problems. Thus teachers can quickly assess which students have mastered skills, along with each student's engagement, affect and motivation, see Fig. 3. In a typical math class, teachers might invite students to discuss the hardest problems in front of the class, projecting the teacher tools in front of the class from the teacher's computer. Other teachers prefer to print booklets that include the hardest problems and have students work in small groups to solve each problem with the aid of a teacher and other helpers.

Teacher tools become a selling point for teachers, supporting them to analyze student strengths and difficulties in specific math skills, currently mapped to Common Core State Standards Initiative (CC).<sup>4</sup> Teacher tools provide a precise assessment of students' strengths and weaknesses at specific math skills, in a language with which teachers are familiar, highlighting both individual math problems and full areas of knowledge (standards) that appear challenging to students. Since each practice activity is internally mapped to standard-based math skills and mastery levels are computed and updated instantly depending on student success at problems, teachers can view an estimate of students' abilities, and how their knowledge has evolved and developed as they used the system. Teachers can log in to the tools to see the hardest problems with which a class is struggling, or the math skill in which students are weakest/strongest. Teachers, and potentially other stakeholders, can receive regular emails about individual and class group progress.

## Performance of Students Using Wayang in the Classroom

Wayang Outpost has been used in middle and high schools as part of regular math classes since 2004, before statewide-standardized test exams. Typically, when the software was used, there were experimental and control conditions to assess the utility and/or impact of various models and variations of the software or pedagogical interventions. During years 2004 through 2006 we carried out studies in an urban school in Western Massachusetts, and in 2012 we carried out in suburban schools in Western Massachusetts. In these studies, we did not always have access to students' MCAS

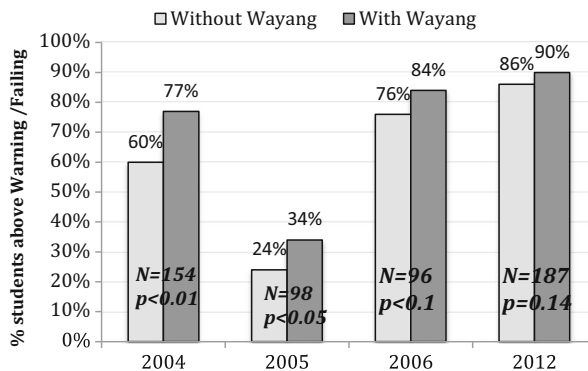
<sup>4</sup> Common Core State Standards Initiative is an effort to ensure that all students in public schools in America acquire the skills and knowledge they need to achieve success at college or in the workforce. This initiative was developed at the state level by leaders concerned about the lack of national standards for the education of children. Nearly all the states in the USA have voluntarily adopted the standards, see <http://www.corestandards.org/Math/>



**Fig. 3** One of the teacher reports highlights math skills (soon Common Core standards) that students in the class found challenging (*yellow*). When clicking on each skill (column 1), a new report is launched that highlights which individual problems involving that math skill were more or less challenging for students in this class

scores; the results shown in Fig. 4 correspond to all the data from Massachusetts's standardized tests (MCAS) that were available to us since 2004 that included matched control groups.

Empirical evaluations of Wayang have repeatedly shown it to be beneficial both for short-term learning (measured by pre to posttest improvement in researcher-created tests) and for longer-term retention (evidenced by state-based standardized test scores, 1 to 4 weeks later). Fig. 4 shows aggregate results since 2004, and highlights that students had higher standardized tests scores at the end of the year after using Wayang Outpost as compared to students in a control condition who did not use the software. We



**Fig. 4** Massachusetts Statewide Standardized Test (MCAS) passing rates for experimental groups (using Wayang, dark grey) and control groups (in regular math class, light grey), within the same school, same grade and same teachers. Passing rates include several ratings above warning/failing

discourage the reader from making comparisons across years, as populations of students and schools change quite drastically from year to year, as well as total number of students and student ratio between experimental and control groups. In particular, our studies in the years 2004 through 2006 involved schools in urban low achieving educational settings, while the larger 2012 study involved a rural high achieving school with better prepared teachers and students.

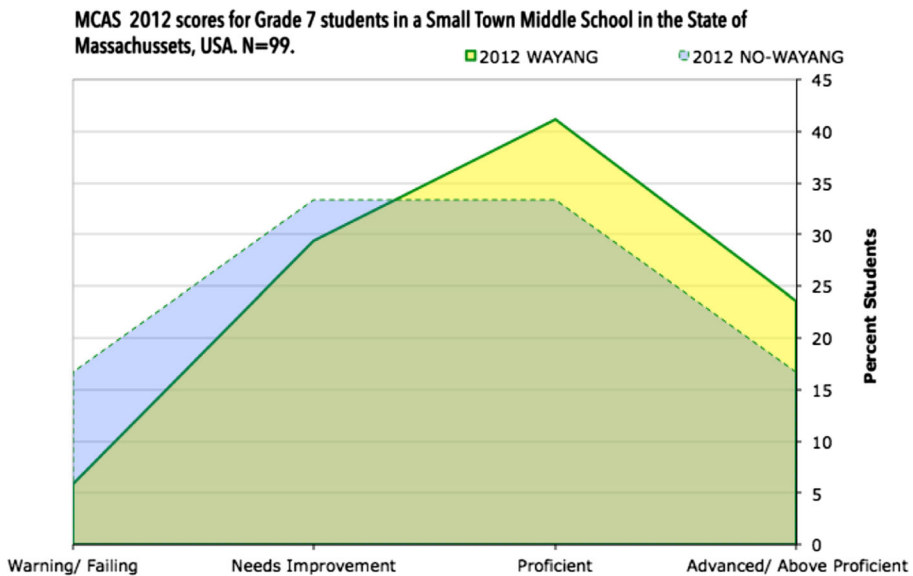
We will look in detail into a controlled study from year 2012 that involved 198 seventh and eighth grade students in a single (average to high achieving) school in Massachusetts. Half of the math classes were assigned to a control condition and half were assigned to an experimental<sup>5</sup> condition, with the experimental group using Wayang Outpost about once a month throughout the full school year. MCAS results at the end of the year revealed a near significant trend in MCAS scores for 7th grade students using Wayang compared to their counterparts not using Wayang (Mean Wayang ( $N=34$ )=244.82; Mean Control ( $N=60$ )=239.50;  $F(1, 93)=2.2$ ,  $p=0.14$ ). As can be seen in Fig. 5, which plots the percent of students in each MCAS scoring category (warning/failing, needs improvement, proficient or advanced), the 7th graders using Wayang shift their performance towards the advanced category compared to the control condition: 12 % more students actually passed the MCAS test than the control 7th graders (leftmost side, warning/failing) and 7 % more experimental students reached the advanced level (rightmost side, advanced/above proficient).

As this was a high achieving school, we analyzed the frequency of both seventh and eighth graders at reaching the advanced (beyond proficient) level. Fifteen percent (15 %) of students in the non-Wayang condition reached the “advanced/above proficient” level, while twenty-one (21 %) of students using Wayang Outpost reached the advanced level. This shift into the advanced MCAS category considering both 7th and 8th grade students approached significance (Chi Square (186,1),  $p=0.17$ ), with Wayang 8th graders having significantly more students in the proficient or advanced category than non-Wayang 8th graders (Chi Square (92,1),  $p=0.05$ ).

Further analyses of the Measures of Academic Progress (MAP)<sup>6</sup> test scores, which was given at the beginning and end of the year to measure growth, indicated a significant difference in student gains from the beginning to the end of the year for one specific area of the mathematics curriculum, the “patterns and algebra” section of MAP, between the control and Wayang groups ( $p<0.01^*$ , effect size *Cohen's d*=0.38). This result was obtained using both seventh and eighth grade data. A more detailed analysis of eighth grade students alone showed largest gains in MAP-Algebra/Patterns than other areas, with an effect size of  $d=0.77$ . Unlike 7th grade teachers, eighth grade teachers had tailored Wayang Outpost's content to mostly contain Knowledge Units related to Algebra, with these eighth grade students seeing most Wayang Outpost problems in this area, for a total of 6109 algebra-related problems seen by individual

<sup>5</sup> There was one extra 8th grade class in the experimental Wayang condition, and one extra 7th grade class in the control condition, resulting in an uneven split across grades: there were more 8th graders using Wayang (MCAS  $N=65$ ) than no-Wayang (MCAS  $N=28$ ), and more 7th graders in the no-Wayang control condition (MCAS  $N=60$ ) than the Wayang condition (MCAS  $N=34$ ).

<sup>6</sup> Measures of Academic Progress (MAP) assessments are designed to provide actionable data about where each child is on their unique learning path. Administered online or locally, MAP helps teachers pinpoint—to the goal—strand level—where students are ready to advance, and where they need help. Longitudinal data from MAP can be used to analyze program impact, see. <http://www.nwea.org/node/98>



**Fig. 5** Area chart comparison of performance for a 7th grade of students on the Massachusetts Comprehensive Assessment System (MCAS), for students using vs. not using Wayang Outpost. Students represented by the yellow/green polygon used Wayang Outpost and students represented by the blue polygon did not use the tutor. Distribution of students using Wayang Outpost shifts to the right indicating that more students passed the exam and received a grade of “proficient” or “advanced” when using Wayang Outpost. Groups of students were matched in terms of teacher of seventh grade students

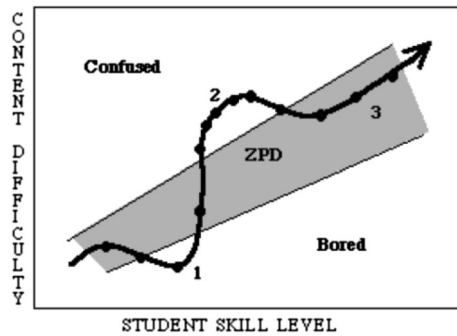
students (72 % of the total math problems seen by eighth graders were of this kind) in “Expressions with Variables, Univariate Equations, Inequalities, and Linear Functions and Relationships”. This provided further evidence of Wayang Outpost’s capability at improving student achievement, as students improved most in those areas of mathematics on which teachers had pre-set Wayang to focus.

### Cognitive Scaffolds and Interventions

We attribute part of the success of Wayang Outpost at improving student performance during the last decade to several components that target student cognition: personalization of the difficulty of selected mathematics problems; provision of multimedia scaffolding and support; worked-out animated examples and video tutorials; and training for retrieval of basic arithmetic skills. This section describes a subset of these components along with their design and evaluation.

#### Adapting Curriculum to Individual Learning Needs

The first component in the Wayang Tutor that we believe led to improved student learning and behavior is an algorithm that aims to maintain students within the zone of proximal development, see Fig. 6. Thus, students who perform well on problems of average difficulty receive harder problems and those who struggle receive easier problems. This novel and flexible framework takes into account: (a) a student’s recent



**Fig. 6** The tutor attempts to maintain the student within a “zone of proximal development” (from Murray and Arroyo 2002). It adapts the curriculum to individual learning needs

performance as well as level of effort exerted; (b) the inherent difficulty of each math problem item estimated from past students’ log files; (c) limitations of content availability; and (d) realistic scenarios of classroom management, such as the time that teachers want students to spend on each topic (e.g., minimum number of problems to be seen in that topic, maximum time to be spent reviewing a topic, etc.) (Arroyo et al. 2010b).

This approach, referred to as “effort-based tutoring” (EBT), adapts problem selection depending on the effort exerted by a student on a practice activity based on three different dimensions of student behavior: attempts to solve a problem, help requested, and time to answer (see Table 1). The algorithm defines an *expected value* of behavior  $E$  for each problem, based on student-problem interactions over several years, across students ( $E(I_i)$ ,  $E(H_i)$ ,  $E(T_i)$  for  $i = 1 \dots p$  where  $p$  is the total number of problem items and  $I$ ,  $H$ , and  $T$  are incorrect responses, hints, and time, respectively). In fact, it defines a region of expected behavior, due to two delta values for each  $E(I_i)$ ,  $E(H_i)$  and  $E(T_i)$ , for a total of six delta values for each problem  $p_i$ , which represent a fraction of the standard deviation, regulated by two parameters,  $\theta_{LOW}$  and  $\theta_{HIGH}$  in the interval  $[0, 1]$ . For example, if  $\theta_{LOW} = 1/4$  and  $\theta_{HIGH} = 1/2$ , then  $\delta_{IL} = \theta_{LOW}SD(I_i) = SD(I_i)/4$  (a fourth of the standard deviation of  $I_i$ ) and  $\delta_{IH} = SD(I_i)$   $\theta_{HIGH} = SD(I_i)/2$ , half of the standard deviation of  $I_i$ .  $\theta_{LOW}$  and  $\theta_{HIGH}$  are the same for all problems in the system.

These values help define “a region of expected behavior” for a practice item within the tutoring system. The reason for having separate deltas for the lower and higher side of the distribution is because these distributions of student behavior are not normally distributed, and instead skewed towards zero. Note that the notation for  $\delta$  values has been simplified (e.g. should really be as it refers to an individual practice item  $p_i$ ).

The left side of Table 1 shows the estimated student state (e.g., mastery without effort, hint avoidance and high effort, etc.), that represents a student’s most likely state of mind (cognitive, metacognitive, motivational) while approaching a new problem, compared to the problem solving behavior for the whole population of students. Thus, the tutor interprets the reason for specific student behaviors (column 4) through comparison with average student behaviors (i.e., expected values), for timing, errors and help received (columns 1, 2, 3). The right side of Table 1 indicates the action taken by Wayang Outpost based on the inferred reason (column 4). One significant benefit of having a pedagogical model in which decisions are made based on orthogonal axes of behavior (correctness, hints, and time, columns 1, 2, 3) is that such decisions can help researchers and the software itself to discern between cognitive, metacognitive and

motivational states. Thus, based on student behavior (left) corresponding to help abuse and disengagement (affective and metacognitive) as well as lack of mastery or knowledge (cognitive), the algorithm indicates which pedagogical moves should be made next by the tutor, in terms of both content difficulty and other actions such as pedagogical agent moves. Thus, Wayang's main pedagogical decision-making method integrates cognitive, metacognitive and affective factors before making a teaching decision based on student learning needs.

Wayang Outpost's pedagogical decisions concern not just content selection, but also affective and metacognitive feedback. Note that student disengagement or low engagement (e.g., Table 1, rows 3 and 5) will result in reduced problem difficulty, based on the assumption that if a student is not working hard enough on the current problem, she/he probably won't work hard on a similar or harder problem. If this might appear to be a somewhat simplistic tutor response, it is important to note that it will be followed by an intervention with a learning companion, an animated digital character (see "[Animated Affective Learning Companions](#)" section) that speaks to the student deemphasizing the importance of immediate success, trying to encourage the student to exert more effort solve problems. These cognitive and motivational scaffolds are designed to increase the likelihood that students will engage further in the next problem. Further details about how the level of challenge is adjusted in order to increase or decrease the difficulty of the upcoming math problems can be found in (Arroyo et al. 2010b). A small randomized controlled study provided evidence that the adaptive problem selection results in improved learning, compared with a control condition where problems were randomly selected (Arroyo et al. 2010b) within the knowledge unit.

The effort-based tutoring (EBT) approach is different in many ways from the more traditional Bayesian knowledge tracing, BKT (Corbett and Anderson 1995; Wang and Heffernan 2013), as explained next. This article does not claim that a non-Bayesian approach to student modeling is superior compared to a Bayesian approach, as further studies would be needed to make such claim. In fact, we have started using hidden Markov models to model engagement and knowledge in parallel (Arroyo et al. 2014; Johns and Woolf 2006) to take a Bayesian approach towards modeling a variety of cognitive, metacognitive, and affective states. In addition, the current Wayang Outpost uses an estimation of knowledge mastery in the traditional BKT sense at the topic/knowledge unit level, and uses it as one factor to decide to move on and across knowledge units; by contrast, EBT is used to adjust content difficulty and other decisions *within* a knowledge unit (e.g., a knowledge unit such as "decimals and percent"). Still, the authors believe that using EBT to make decisions within the knowledge unit affords several improvements to traditional knowledge tracing or other theories such as Item Response Theory (IRT) (Rasch 1960) used in computer-adaptive testing. Specifically:

- a) EBT models a combination of the cognitive, affective and metacognitive states of students as they interact with the tutoring system, based on combinations of timing, correctness and hint requests. Traditional BKT models only the cognitive component of students' knowledge;
- b) EBT captures the fact that some problems are inherently harder or easier, regardless of the mathematical procedure required to solve each problem and depending on operand size and other characteristics of the problem such as spatial distribution.



Traditional BKT models assume that problem difficulty is related to student knowledge (where less knowledge means greater difficulty). For instance, the two problems shown in Fig. 2 are exactly the same from a procedural perspective—they require exactly the same steps for the problem to be solved, however, students are more familiar with horizontal parallel lines, thus the item on the bottom of Fig. 2 is harder for students in general. The EBT approach can capture differences in problem/item difficulty without having to explicitly model the knowledge that might account for these difficulties;

- c) EBT models item difficulty in terms of a continuous range of correctness (attempts), and takes into account timing or help received by the student. Traditional IRT systems do model item difficulty and do capture items that might contribute differentially towards evidence leading towards knowledge, but they do not do it in terms of a continuous range of different student behaviors, beyond correct/incorrect;
- d) EBT combines modeling with an optimization mechanism (the *acting* component), which depends on very recent student performance, continuously searching for the ideal content for an ever moving target—a student’s changing knowledge, affective and metacognitive state. Traditional BKT and IRT are methods used to model student knowledge states (modeling towards prediction) and their relationship to correctness to respond, and the Cognitive Mastery approach uses BKT and selects problems for not-yet mastered skills on an individual basis, until the student reaches mastery on all skills, about .95 probability of knowing;
- e) EBT is simple, scales up easily as new material is added to the system; it also learns from students as new data is input; and easily integrates different activity formats (e.g., short answer vs. multiple choice vs. other forms of responding such as clickable/draggable answers) within the same system.

### Training Students In Basic Cognitive Skills

The second component in Wayang Outpost that has shown improved student learning is math fluency training (Arroyo et al. 2011a). This involves training very basic arithmetic skills such as addition, subtraction, multiplication and division of single and double digits numbers, focusing not only on accuracy (which might be at ceiling performance) but especially on speed to respond. Based on an information-processing model of the brain (Baddeley 1986), this approach is also called math facts retrieval (MFR) Training, and it attempts to reduce working memory load by automating math fact retrieval from long-term memory into working memory, and by developing automaticity in basic math operations (Tronsky 2005; Tronsky and Royer 2003). As will be described shortly, our experiments have shown improved learning among middle school children on math-standardized tests, when sessions of Wayang Outpost are preceded with MFR training, as compared to Wayang alone, or MFR training alone.

The underlying cognitive theory is based on retrieval from long-term memory as an important skill in mathematics activities, since problem solving takes place in a cognitive system constrained by a limited capacity of working memory. Many students have difficulty with mathematics problems, in part because they are slow and/or inaccurate in retrieval of simple math facts from long-term memory. Training in the speed and accuracy of very basic math skills has been shown to be especially effective



for students with learning disabilities, who may show number processing inefficiencies (Royer and Walles 2007).

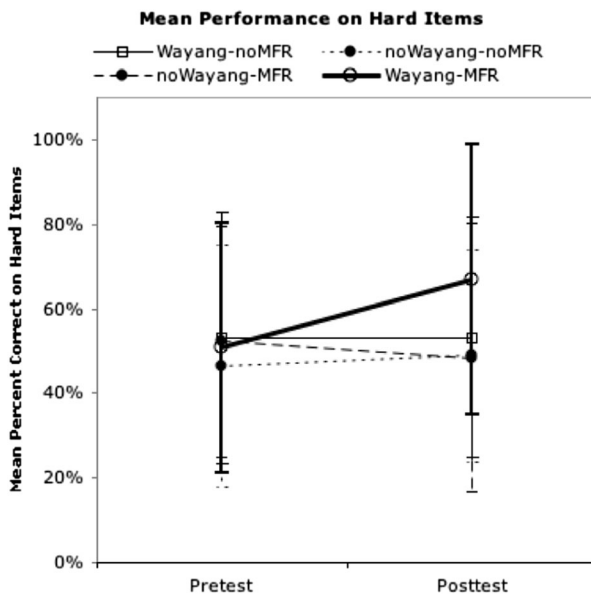
The math fact retrieval software, *MathSuccess*,<sup>7</sup> is an off the shelf product that is independent of Wayang. Originally developed for reading and dyslexia and then extended to mathematics fluency, the math facts retrieval software provides both *training* and *assessment* that blends an auditory/verbal with a visual/spatial component of fluency. In the training phase, students study pages of math facts (e.g., two operand addition/subtraction/multiplication/division of at most two digit numbers). Students click on each item to hear the answer and “study” individual math facts, as if they are studying vocabulary words. During the assessment phase, students are tested for their accuracy and speed (recorded at millisecond resolution). Students speak the answer aloud and immediately hit the space bar, after which the correct answer is *spoken back* to the student. Students self-score themselves with the right/left buttons of the mouse, and at the end of the assessment session, they can see a graph chart that shows their progress (in speed and accuracy) at retrieving easy (already mastered) math facts compared to the previous assessment session. Students normally become faster at retrieving math facts as they work on more pages. Progress charts show them how much faster and accurate they are getting, which serves as a motivation to “go for another round”.

A time-controlled study involving 250 middle school students in a Massachusetts school produced encouraging results for the combination of MFR and Wayang Outpost (Arroyo et al. 2011a). Students were randomly assigned to one of four conditions: (1) use of Wayang Tutor alone (Wayang-noMFR); (2) use of Wayang Tutor after working on the MFR Training software for 15 min (Wayang-MFR); (3) use of the MFR Training software (noWayang-MFR) and then use of other modules and web sites (e.g., National Library of Virtual Manipulatives) that did not explicitly tutor mathematics; and (4) classroom instruction instead of software instruction or use of math web sites (noWayang-noMFR). Students used the technology instead of math class.

There were significant effects for Wayang and the combination of Wayang and MFR training, indicating that the Wayang-MFR group did better than the Wayang-noMFR group. The effect size for Wayang vs. no-Wayang groups (Cohen’s *d*) was 0.39, and the group with the highest scores at posttest time was the Wayang-MFR group that received both Wayang and MFR training (Fig. 7; for details, see Arroyo et al. (2011a)). As expected, students with MFR training became faster at responding to easy arithmetic items, see Fig. 8.

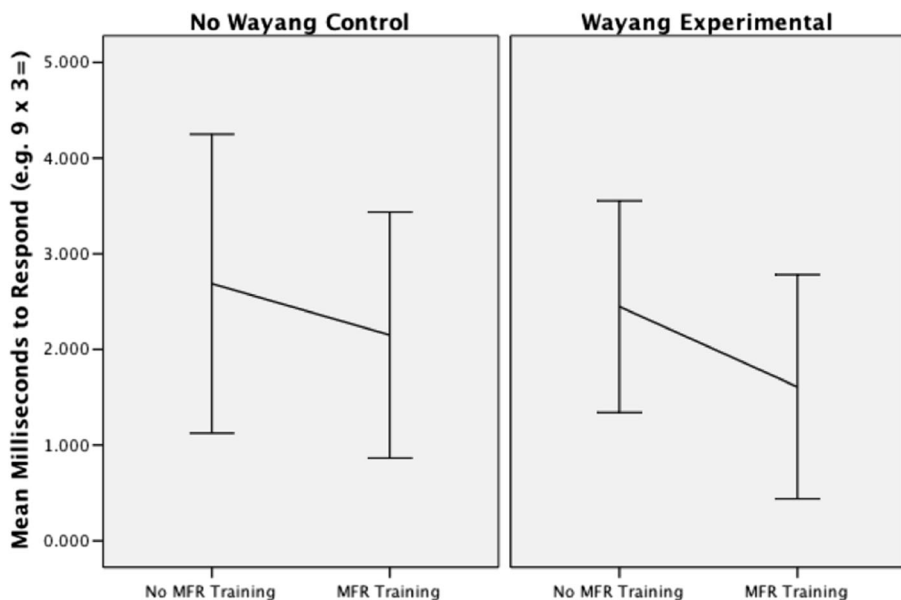
One interpretation of these results is that increased math fluency, resulting from MFR training, frees up cognitive resources that are essential to solve challenging math problems. Easy problems do not require as much working memory, thus MFR training does not have an impact on performance for these items. This evidence that suggests that training provided by intelligent tutors for mathematics can be enhanced if students also receive training in *speed* of foundational skill activities – even if those skills are at ceiling accuracy before tutoring starts (i.e., mastered at pretest time). A combination of fluency training (the speed with which students either retrieve or calculate answers) and

<sup>7</sup> <http://www.mathsuccesslab.com>



**Fig. 7** Mean improvement (and standard deviations) on hardest items of the math pre/posttest. The *thick line* represents students who received both the Wayang Tutor and math facts retrieval training software; all other groups did not really improve on these harder multi-step items

strategic training (approaches to solve specific kinds of problems) should yield higher success at more complex problem solving. Efforts are currently in place to integrate the MFR software into the Wayang Outpost/MathSpring tutor, and research continues to



**Fig. 8** Means and Standard Deviations of milliseconds to respond to basic arithmetic problems. Students who received MFR Training (to the right of each chart) became faster at solving simple arithmetic problems as compared to students who did not receive MFR (to the left of each chart)

understand the limits of which and how pre-requisite math skills should be trained towards automaticity.

### Metacognitive Scaffolds and Interventions

Wayang Outpost also includes components that target metacognition, by which we mean cognitive resources and mechanisms that help students to regulate their own learning. We evaluated several interventions that involve 1) open student models that scaffold the self-regulatory process, encouraging reflection and informed choice at key moments of boredom, 2) progress charts and tips that encourage good study habits, and 3) interventions supporting help-seeking behavior in order to improve self-monitoring and evaluation. These components will be discussed later in this section, after we present the theoretical foundations for this work.

In general, support for self-regulation in Wayang Outpost is based on several models of learning (Butler and Winne 1995; Greene and Azevedo 2007). These models define the learning process as a flow model where each individual uses strategies to produce an outcome that is then subjected to external feedback or revision based on an internal monitoring loop within the cognitive system. These researchers propose that learning occurs in the following phases: task definition, goal setting and planning, studying tactics and strategies, carrying out a plan to generate a product, comparing outcomes to standards, and adaptations to metacognition. Monitoring is an important and necessary process while solving a problem, while evaluation occurs at the outcome or product level. Based on Butler and Winne's model, and following work by Azevedo et al. (2005), we hypothesize that an optimal tutoring system needs to help students analyze the learning situation, set meaningful learning goals, determine which strategies to use, assess whether the strategies are effective, and evaluate their own emerging understanding of the topic.

Another model that inspired our work on metacognitive scaffolding is that of Zimmerman and Moylan (2009), which is a socio-cognitive model of self-regulation that adds a motivational/affective component to self-regulation. In their model, students loop through three cyclical phases: forethought, performance and self-reflection. The forethought phase refers to motivational/affective processes that precede efforts to learn, and which influence students' predisposition to start or continue the learning process. Performance involves processes that occur during studying and/or problem solving and impact concentration and outcomes (including monitoring during problem-solving execution). The self-reflection phase involves processes that follow problem solving or studying efforts, with a focus on a learner's reactions to the experience (including self-evaluation and self-judgment). These self-reflections, in turn, influence forethought regarding subsequent learning efforts, which completes the self-regulatory cycle.

Dignath et al. (2008) extend these ideas in a large meta-analysis of self-regulatory training programs, which highlights that self-regulation training programs have been effective at improving primary school students' academic performance (cognitive outcomes), strategy awareness and use (metacognitive outcomes), and motivation (affective outcomes). Given these models and findings it is reasonable to believe that intelligent adaptive learning environments should not only support self-regulation but also include components that explicitly teach self-regulation. The distinction between

teaching and supporting metacognition is that for teaching metacognition, the goal is to improve students' metacognitive behavior, even after the metacognitive scaffolds are removed. In contrast, supporting metacognitive means that the goal is to improve learning while the metacognitive scaffolds are in place.

We conjecture that emotions arise throughout all phases of the metacognitive process. Pekrun introduced the control-value theory of achievement emotions (Pekrun 2006; Pekrun et al. 2007), in which emotions are classified as prospective (future expectancy of performance outcomes); retrospective (regarding past performance evaluations) and activity-based emotion (experienced during performance/studying). Combining theories from Pekrun (2006) and Zimmerman and Moylan (2009), we propose that prospective emotions are more likely to occur in the forethought phase, activity emotions in the performance phase, and retrospective emotions in the self-evaluation phase –this is still an issue that deserves further investigation. In addition, negative valence emotions during any of these phases can make students disengaged, degrade performance and eventually make students quit out of the self-regulatory loop (give up).

Based on these foundations, we incorporated several new components into the Wayang Outpost Tutor to support student metacognition. Our focus is on self-regulation of students' learning in order to address disengagement and other non-optimal student experiences observed in student-tutor interactions that we consider are in part consequences of a degraded self-regulatory cycle.

### Open Student Models to Enhance Self-Regulation

The first component in the Wayang Tutor that targeted metacognition was an open student model called student progress pages (SPP), see Fig. 9. Promising research into



**Fig. 9** The open student model in Wayang is called the Student Progress Page (SPP). It encourages students to reflect on their progress for each topic (column 1). The plant (column 2) demonstrates the tutor's assessment of student effort, while the mastery bar (column 3) records presumed knowledge (according to Bayesian Knowledge Tracing). The tutor comments on its assessment of the student's behavior (column 4) and offers students the choice to continue, review or challenge themselves and make informed decisions about future choices (column 5)

the relation between metacognition, student learning (Zimmerman 2000), and supporting students' self-regulation (Aleven et al. 2010; Roll et al. 2007) shows that open student models promote metacognitive activities by encouraging students to reflect on their progress, to take greater control and responsibility over their learning, and to increase their trust in an adaptive environment through increased transparency (Bull 2012; Kay 2012; Long and Aleven 2013; Mathews et al. 2012; Thomson and Mitrovic 2010). The open student model in Wayang allows students to inspect their progress and the tutor's assessment of their work, see Fig. 9 (Rai et al. 2013). Students use the open student model to make choices about which topics and difficulty level to work on next. They also use it to monitor their own performance, to receive feedback about their progress, to reflect on their learning, and to make informed choices. Students can explicitly visualize their effort excerpted in problems via a plant that blooms and gives peppers; the effort displayed is problem-based and represented based on total amount of effort scenarios as shown in Table 1, i.e. being engaged with a problem, spending time in it and asking for hints when mistakes are made. Currently, students cannot modify the student model, or dispute the assessment of their own knowledge.

The student progress page lists mathematics topics (rows) and encourages students to reflect on the tutor's assessment of their effort and knowledge (Fig. 9, columns 2–3 respectively) supporting students in the self-assessment stage. The SPP also supports students to stop and set new goals through buttons in the last column that allow students to choose to continue, review or challenge themselves and to make informed decisions about future choices. Students might also switch to a topic that they might be weaker in and need further "growing," column 1. The SPP is designed to encourage many of the behaviors that have been predicted to be beneficial with respect to open student models (Bull 2012; Bull et al. 2012; Kay 2012; Mathews et al. 2012), including encouraging students to reflect on their progress, supporting them to take greater control and responsibility over their learning, and increasing their trust in the environment.

In an evaluation of this approach of how a "metacognitive intervention" might cross over and produce an improved affective outcome, one experimental condition invited students to use the SPP when it detected a deactivating negative affective state (boredom or lack of excitement). Students in the control condition had the same SPP available via a button, but the tutor did not suggest its usage at any point of time. The purpose of the experiment was to compare how students behaved using the tutor, what emotions they reported, and what emotions our affect detectors predicted they experienced, when being encouraged or not to visit the student progress page to encourage evaluation and goal setting (continuing, reviewing, challenging themselves, switching topics, etc.).

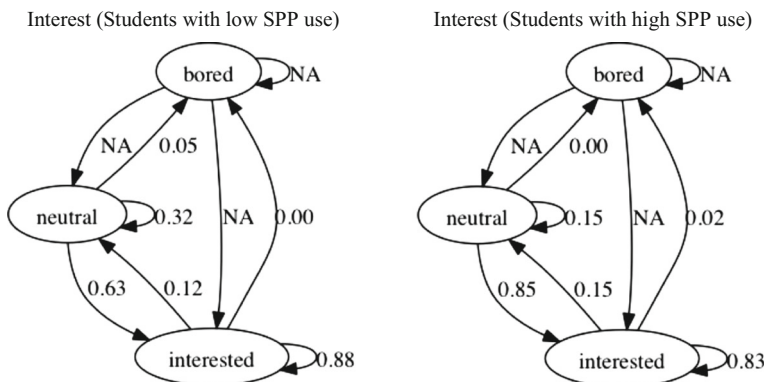
The decision to offer to see the SPP was based on students' actual self-reports of their emotion. Unfortunately, according to student self-reports, students were almost never bored in this experiment, so Wayang almost never suggested that they use the SPP. As a result, the total amount of use of the SPP was not reliably different between the experimental and control groups of students. Given this, we analyzed differences between students who had low vs. high SPP usage based on a median split of SPP access (comparing students who used the student progress page more frequently vs. less frequently). We created affect detectors specific for this data set using a variety of

features and techniques that we have used in the past (Wixon et al. 2014), which allowed us to understand how students felt during each and every math problem. We then used Markov chain models for our analyses, path models that show how students transition between emotional states from math problem to math problem, see Fig. 10.

These path models show that students with high-access to the SPP have a higher probability to transition from a ‘neutral’ emotional state to a state of ‘interested’, instead of remaining in a neutral state or becoming bored, Fig. 10, right. On the other hand, students who demonstrated low usage of the SPP (Fig. 10, left) were more likely to transition from a neutral state to one of boredom (+.05) and were less likely to transition from a neutral state to one of interest than the student with high SPP usage. Unfortunately, this might be an effect similar to the “rich get richer”, meaning that students that already had good metacognitive behavior benefitted most from the metacognitive feedback and the system as a whole. These studies need to be repeated to verify the kinds of metacognitive and affective benefits the student progress page can produce.

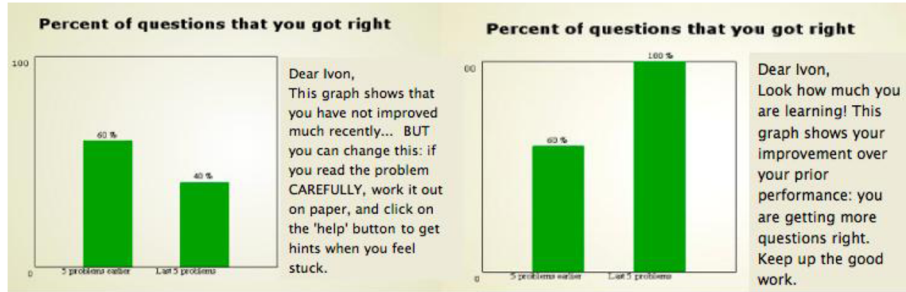
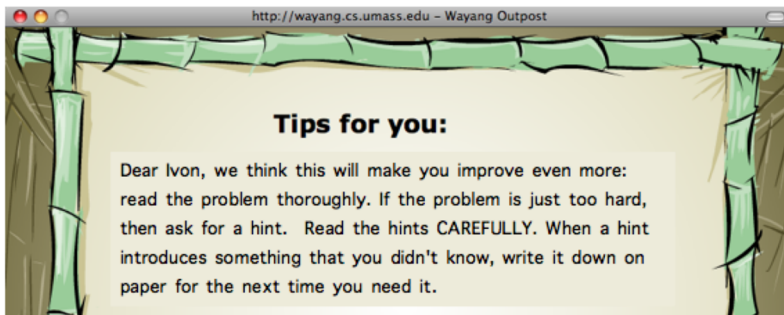
### Progress Reminders to Enhance Self-Regulation

Another component in Wayang Outpost that targeted self-regulation was a suite of reminders of student progress, including charts and tips. This intervention corresponded to two intervention screens that appeared after fixed intervals every six problems (i.e., after students clicked the ‘next problem’ button on the sixth problem). The intervention screens were shown to all students, but their contents were driven by the student’s behavior within the system. Students received either 1) a progress performance chart with an accompanying message, see Fig. 11.a (negative/positive bars depending on recent and past performance); or 2) a tip (message) that encouraged productive learning behavior, see Fig. 11.b (“Dear [student’s name], We think this will make you improve even more: Read the problem thoroughly. If the problem is just too hard, then ask for a hint. . .”). Students were addressed by their first name in the messages accompanying



**Fig. 10** Markov chain models represent student transition probabilities from one learning activity to another, between states of boredom and interest. The overall likelihood for students remaining interested was 83 % for students who accessed the SPP most (*right, bottom*), a bit lower than the 88 % for students in the low-access group (*left, bottom*). However, students with high SPP use had a higher likelihood to transition from neutral to interested (0.85) than did students in the control condition (0.63), a 22 % difference. Students with high SPP use (*right*) were less likely to remain in the neutral state (0.15) than students with lower SPP use (0.32)



**a****b**

**Fig. 11** **a.** Progress Charts in Wayang show students the accuracy of their answers. **b.** Tips in Wayang encourage good learning habits

both charts and tips. Whether a student saw a progress chart or a tip and which one, was randomly decided.

Eighty-eight students from four different classes (10th grade students and some 11th grade students) from an urban-area school in Massachusetts used Wayang Outpost for 1 week during four class periods. Students in the experimental condition used Wayang with progress charts and tips. One control group used a version of Wayang that lacked progress charts and tips. A second control group (called no-tutor control) consisted of matched classes of students who did not use the Wayang Outpost tutor at all. These students were of the same grade level, equivalent proficiency level, and taught by the same teachers as the students in the other conditions. Students were randomly assigned to either an experimental control conditions, and the experiment controlled for time. Further details on this study can be found in (Arroyo et al. 2007).

Mathematics performance was evaluated with pre- and posttests; these instruments also included questions on mastery learning orientation and students' liking of mathematics (Mueller and Dweck 1998; Wigfield and Karpachian 1991). The post-tutor survey also inquired about how human-like the tutor was. Another measure used corresponded to disengagement (a form of gaming, assessed via an automated gaming detector specified in Johns and Woolf (2006)), which consisted of estimations in relation to each problem within the tutoring sessions. If the interventions were effective, students' gaming behaviors would be reduced, as found for instance in (Baker et al. 2008).

The interventions did influence student behavior. In particular, students in the experimental group displayed significantly fewer cases of quick-guessed answers to



problems (a form of gaming defined as rushing to answer in less than 4 s) in the problems following the intervention: 12 % quick-guessed in the experimental condition vs. 18 % in the control condition (based on analysis of every sixth problem), a significant difference of 6 % more guessing in the control condition ( $p < 0.01$ ). Also progress tips and charts increased student focus in the math problem immediately after the intervention, e.g., improved time spent on a problem (increased focus) was displayed in the problem immediately following the intervention.

The interventions also influenced student learning. The raw learning gain (posttest-pretest) for the experimental group (ProgressTips) was 9 % while students in the Tutor Control group showed no improvement. Table 2 shows the pre- and posttest scores for no-tutor control (students who did not use Wayang - top row); tutor control group (students who used Wayang without the interventions - middle row) and the intervention group (students who used Wayang with progress tips - bottom row). We analyzed the difference in learning gains between the two groups (progress tips tutor and tutor-control). A significant difference between the Tutor-Control and ProgressTip Tutor groups ( $p < 0.05$ ) indicates that the interventions-enhanced group learned more. As shown in Table 2, students who received ProgressTips passed the state standard exam more frequently, 92 % vs. 79 % (or 76 % for students who did not use any tutor), see Arroyo et al. (2007) for further details.

Note that, across all conditions, the learning gains are smaller than what we had observed in previous studies with Wayang Outpost (the latter were typically 15 % in about the same amount of time). We think this may be due to the fact that, in prior studies, Wayang provided tailored problem selection based on student needs, following the problem selection decisions in Table 1. During this study Wayang used a fixed sequence of problems. Eliminating the adaptive sequencing was done to reduce variation across conditions and across engaged/disengaged students.

Further unpublished results are described next. We analyzed how the target interventions impact student disengagement behaviors (gaming) and learning. Gaming is defined as “attempting to succeed in an educational environment by exploiting properties of the system rather than by learning the material and trying to use that knowledge to answer correctly” (Baker, R. S. J. d et al. 2006). Gaming can be a sign of poor self-regulation and in some instances gaming harms learning. However, as Baker et al. (2008) have pointed out, gaming is not always detrimental to learning, as students might have a “gaming style” and can carry out non-harmful gaming that does not impact their learning. Thus, we split students at their median gaming rate to classify them as low vs. high gaming students (where gaming was identified as quick-guessing on a problem, hint-abusing –reaching the last hint that gave the answer quickly– or skipping problems without taking any action in the tutoring sessions). We then made a

**Table 2** Students in the experimental group (last row) received tips and charts every 6 problems. Means and standard deviations in performance measures before and after tutoring for the three groups

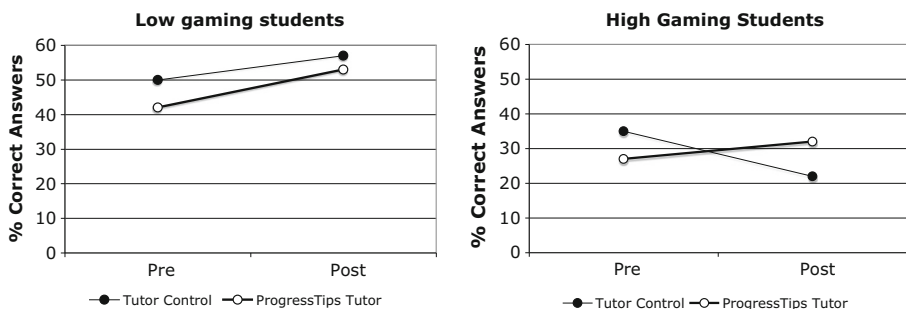
Group	Math Pretest	Math Posttest	Passing Rate in State Standard Exam
No Tutor Control			76 % ( $N=38$ )
Tutor Control	40 % (20) ( $N=40$ )	40 % (28)* ( $N=40$ )	79 % ( $N=34$ )
ProgressTips Tutor	33 % (19) ( $N=36$ )	42 % (22)* ( $N=36$ )	92 % ( $N=24$ )

comparison depending on gaming style. Would high gaming students also be harmful gamers? Would progress charts with Charts and Tips benefit high gaming students or low gaming students?

In general, gaming negatively impacted learning. High gamers' performance improved only very slightly (Fig. 12, left) whereas low gamers' performance improved by about 10 % from pre to posttest (Fig. 12, right). However, these Progress Reports had a differential impact for high gaming students. When high-gamer students did not receive charts and tips their performance decreased from pretest to posttest (Fig. 12, right, black circle). On the other hand, high gaming students who received Progress Charts and Tips (Fig. 12, right, clear circle) improved from pre- to posttest. These students also became less-harmful gamers than students in the control group. High gamers had significantly higher posttest score overall in the Progress Reports group  $F(1,76)=3.4$ ,  $p=0.02$  compared to the standard Tutor control group, after accounting for pretest as a covariate. These results are similar to ones obtained with Scooter the Tutor (Baker, R. S. J. d et al. 2006), where specific interventions especially benefitted high gamers. In the case of Wayang Outpost, when high-gamers were shown Progress Reports, their learning rates improved.

Because the experiment was carried out several days before a statewide-standardized test exam in Massachusetts (MCAS), we collected standardized scores for students who took the exam (only a subset of ~30 students in the experiment), and also collected scores for a group of students (same level and same teachers) who did not use Wayang. Thus, the MCAS acts as a delayed posttest. Students in the ProgressTips tutor group obtained a higher passing rate on the MCAS exam (92 % vs. 79 %) than did their counterparts in the control group, and this difference approached significance (ChiSquare (1,62)=2.4,  $p<0.12$ ). This finding is further evidence that the ProgressTips interventions had an impact on students' mathematics problem-solving ability (note that the comparison of tutor conditions vs. no-tutor condition are in Fig. 4, year 2006).

Last, we measured the affective impact of Progress Charts and Tips. Students in the ProgressTips tutor group agreed more with a variety of statements that attributed a positive human-like qualities to the tutoring system, such as "the Wayang tutor was smart and friendly". They also had significantly higher mastery learning orientation scores, as measured by two survey questions based on ones by Mueller and Dweck (1998), indicating that students who saw Progress Charts and Tips had a greater desire to learn for the sake of learning.



**Fig. 12** High gaming students improve math performance when they receive progress tips and interventions (*left*) but not when they don't receive interventions (*right*)

## Scaffolding Help Seeking Behavior towards Enhancing Self-Regulation

Another component in the Wayang tutor that targeted self-regulation was the teammate effect, which we found supports help-seeking behavior. We believe that an important connection exists among student self-monitoring, self-evaluation and help seeking. Help seeking is an attribute of self-regulation, as described by learning science theories (Butler and Winne 1995; Greene and Azevedo 2007). Help seeking occurs after a student becomes aware of needing help, and is a proactive decision that enables students to continue making progress in the problem solving process.

The decision to seek help is potentially related to students' perceived competence during the self-evaluation and forethought phases. Thus, some students may avoid seeking help, probably in an effort to protect their self-perception of competence. Such avoidance is detrimental to student learning (Aleven et al. 2006). In general, students are challenged by asking for help when needed; often students engage in non-optimal behaviors, such as quick guessing or abusing help, when they search for the answer (Aleven et al. 2005, 2006).

Students' relationship between their view of themselves and the act of help seeking might well depend on students' mindset, e.g., their belief about intelligence being static (fixed) vs. dynamic (can be increased) (Dweck 1999). If a student believes that intelligence can be increased, the student might approach a task relentlessly and seek help when needed; instead, if a student believes that intelligence and math ability are fixed, he or she might feel threatened by the idea of needing help. Thus, it is pertinent to design adaptive help scaffolds within intelligent tutoring systems that teach students how to appropriately seek and utilize help (Aleven et al. 2006), and to promote a better affective relationship toward help seeking in general.

While exploring ways to enhance students' help seeking behaviors, we studied the *teammate effect*, by manipulating the tutor's interface to enhance students' belief that they were working in partnership with the tutor. We hypothesized that students who thought they were working within a team that included Wayang Outpost might manifest more positive help seeking and learning behaviors. This was based on the belief that a basic factor in the relationship between students and the tutoring system is the students' perception of the role of the tutoring system within their learning process: Is it a substitute teacher, a helper, a friend? Do students treat the system as a human or as a tool? Research has shown that by encouraging team relationships between humans and computers, people can be influenced to think that the information from the computer is of better quality and relayed in a more friendly way (Bickmore and Picard 2004; Nass et al. 1996). Moreover, Ogan et al. (2011) further showed that social interactions in a learning environment resulted in greater learning.

We explored these notions by investigating whether students who are encouraged to build a teammate relationship with the computer by manipulating the tutor's interface to suggest that students were collaborating with it. We asked whether students using the teammate approach are motivated to engage in more productive use of the system, with more effective help-seeking behaviors. Specifically, we wondered whether students: 1) would have better learning outcomes when they learned with Wayang Outpost in the role of a teammate and 2) would engage in more effective help-seeking behaviors when they worked as a team to solve problems with the computer. The following is a summary of studies further described in Tai et al. (2013).

Ninety-seven students from a small town middle school in the state of Massachusetts, USA, were randomly assigned to experimental and control conditions. Students in the experimental group were encouraged to relate to the tutor as a teammate via a help button that was labeled “Work Together”. Students in the control condition received a help button that was labeled “Help”. In order to enhance the team relationship, students in the experimental condition also received a prompt at login time saying: “Dear<student’s name>: we encourage you to solve math problems with the tutor character as a teammate. Click on the “work together” button if you don’t know how to solve the problem, so the tutor can help you”. Students in the control group were prompted only with “Dear<student’s name>”, and they were told to ask for help by clicking on the “Help” button if they did not know how to solve problems. Students in both conditions received exactly the same content of step-by-step hints if they asked for help from the tutor. They saw a prompt screen every time they logged in to remind them to ask for help if they did not know how to solve a problem.

Results indicated that the teammate manipulation did not lead to improved learning, but it did impact help-seeking behaviors, with a higher total amount of hints requested by students in the experimental condition as compared to the control condition ( $p < 0.05$ ). While students saw more hints, there was no significant difference between the conditions in the frequency with which students quickly went to the bottom-out hint, a key form of hint abuse.

These results suggest that when students are encouraged to consider the tutor as their teammate, they will ask for more help and thus get more support from the tutor. Students in the experimental condition also had a lower frequency of quick-guessing on problems (quickly attempting an answer with not much time to even read the problem). Students in the experimental group saw many more hints, and apparently they used them in a positive way – as there was no significant difference with regard to hint abuse across conditions (Tai et al. 2013).

These results are consistent with an earlier study that found improved help-seeking behavior but no improvement in students’ learning outcomes (Roll et al. 2011).

## Affective Scaffolds and Interventions

While previous sections described several components that target cognition and metacognition and showed evidence that these components support a variety of learning outcomes, other components in Wayang targeted students’ affective states. Doing so could improve student engagement with the system and, as learning a consequence, because students’ affective *states* and *traits* (e.g., frustration, boredom) can bias the outcome of any learning situation, whether human or computer based. The concept of student motivational states and traits comes from Dweck’s theory of motivation and praise (Dweck 1999, 2002a, b). This theory holds that students who view their intelligence as fixed and immutable (trait-based) tend to shy away from academic challenges, whereas students who believe that intelligence can be increased through effort and persistence (state-based) tend to seek out academic challenges. Students who are praised for their effort are more likely to view intelligence as being malleable, and their self-esteem will remain stable regardless of how hard they have to work to succeed.

Student emotions within a traditional classroom have been described as control or value-oriented (Pekrun 2006; Pekrun et al. 2007, 2010). This control-value theory is based on the premise that student appraisals of control and value of a task are central to the arousal of achievement emotions, including *activity*-related emotions such as enjoyment, frustration, and boredom experienced while learning, as well as *outcome*-related emotions such as joy, hope, pride, anxiety, hopelessness, shame, and anger. Students often use coping strategies, e.g., avoidance, humor, acceptance, and negation, to regulate their emotions in stressful learning situations (Eynde et al. 2007).

Student emotions can be influenced in a variety of ways. For instance, the presence of someone who cares, or at least appears to care, can make students' experiences more personal and help them persist at a task, even a computer task (Burleson and Picard 2007). Moreover, when people are in a certain mood, whether elated or depressed, that mood is often communicated to others; thus a student might register joy or sadness from someone nearby who exhibits those emotions. Feelings are *contagious*: when we register a feeling from someone else, there are signals in our brain that imitate that feeling in our bodies (Hartfield 1994). *Empathic* responses from a teacher or graphic character might work when students do not themselves feel positive about a learning experience (McQuiggan et al. 2008). Thus, a computer persona that appears to enjoy math experiences could transmit these positive feelings to students. The literature suggests that empathic responses *from a teacher* work well in situations where students do not feel positive about the learning experience (Graham and Weiner 1986; Zimmerman 2000).

In general, if computers can model and understand students' emotion in real time, they can begin to *act* upon students' emotional states, encouraging them to use more productive coping strategies. Computers might further attempt to understand the causes of negative affective states, as well as explore the utility of various responses that are not necessarily affective in nature.

The field of AI and Education has made great strides in devising computational models that recognize student emotion (Graesser et al. 2007; Lester et al. 1999; Robison et al. 2009) and is starting to explore how to use that information to respond to student emotion (D'Mello et al. 2010; Rai et al. 2013; Tai et al. 2013). We have created a series of affect detectors, specifically classifiers based on linear regression models that predict four different student emotions from recent student behavior with the tutor. Initially, physical sensors (camera, seat cushion, etc.) were used to predict students' emotions within the software, see Fig. 13, (Arroyo et al. 2009a, 2010b;



**Fig. 13** Four physiologic sensors used to measure student emotion. The emotion component measured (and the physiologic sensor used) included: 1) facial expression (camera); 2) increasing amounts of pressure placed on mice related to increased levels of frustration (mouse with accelerometers), 3) skin conductance (wireless conductance bracelet based on an earlier glove developed at the MIT Media Lab); and 4) elements of a student's posture and activity (chair with pressure sensitive seat cushions and back pads with incorporated accelerometers)

Cooper et al. 2009, 2011). Currently, sensor-free detectors of student emotion have been generated and scaled-up to new schools and populations (Wixon et al. 2014; Baker, R.S.J.d et al. 2012) using information based on recent student behaviors, patterns of behaviors and other trait-based student descriptors.

#### Foundations: Detection and Responding to Student Affect

At the time we started our research in 2004, no comprehensive, validated, theory of emotion existed that addressed learning, explained which emotions are most important in learning, or identified how emotion influenced learning (Picard et al. 2004), though later work by Pekrun solidified that idea that our intuitions were headed in the right direction (Pekrun 2006). At the time, we identified a subset of emotions based on Ekman's analyses of facial expressions that includes joy, anger, surprise, fear, disgust/contempt, and interest (Ekman 1999), with the intention of recognizing these emotions in student behavior and then providing interventions. We added a cognitive component to ground this emotion categorization in an educational setting, resulting in four orthogonal bipolar axes of cognition-affect (Arroyo et al., 2009a), and this resulted in a set of emotions congruent with the control-value theory of achievement emotions (Pekrun 2006). Two of the four axes of emotions were bipolar, in the sense that they had emotions at each end: "confidence/hope...anxiety", "interest...boredom"; the other two were unipolar: "frustrated ... not frustrated", "excitement ... not fun". We tend to refer to these emotions as *confidence*, *interest*, *frustration*, and *excitement*, though the reverse of the first two are also anxiety and boredom.

#### Physiologic Sensors and Interaction Data to Measure Student Emotion

The first component in Wayang that detected student emotion was a suite of physical sensors, see Fig. 13. We developed sensors that captured students' physiological responses while they interacted with Wayang; this data was then combined with information coming from a student's interaction with Wayang (Cooper et al. 2009). The hardware (with the exception of the camera affective facial recognition software that was developed at MIT) was advanced at Arizona State University based on validated instruments and systems first developed by the Affective Computing group at MIT (Picard et al. 2004). To evaluate the utility of the physical sensors for affect detection, we invited high school and university students to interact with Wayang for 4–5 days and outfitted them with all four sensors, see Fig. 13. Wayang iterated through different mathematics topics and problems were chosen adaptively depending on students' ongoing math performance, as specified in previous sections (Arroyo et al. 2010b). We still needed a 'gold standard' of affect, information on how students were feeling, to compare with the sensor data. Thus, every 5 min and after students finished a problem, a screen queried them about their emotions, randomly asking about a single emotion selected from a pre-specified list: "How [interested/excited/confident/frustrated] do you feel right now?" Students choose one of 5 points on a continuum, where the ends were labeled (e.g. I feel anxious ... Very confident) and where "3" corresponded to a neutral value.

Results showed that our sensors in conjunction with the interaction data predicted over 60 % of the variance of students' emotional states. To illustrate, the variables that



predicted *confidence* included Solved? (did the student eventually solve the problem correctly?) and concentratingMax (the maximum probability that the student was “concentrating”) a value provided by the facial expression software.

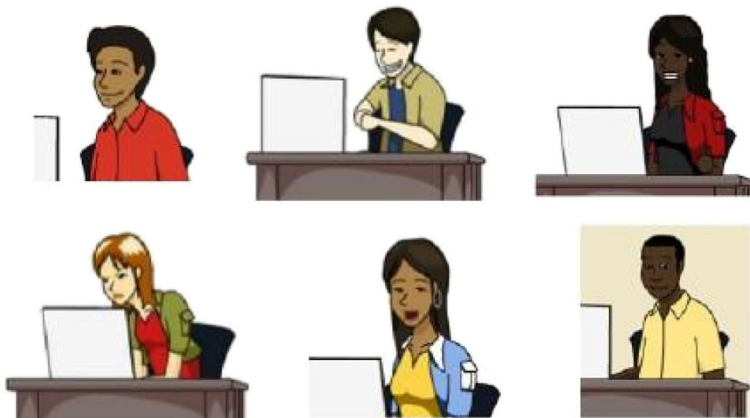
When we analyzed what emotions students said they experienced, we found that these emotions were highly dependent on the tutoring scenario, particularly on indicators of effort in the last problems seen (Arroyo et al. 2009a). These fluctuating student reports were related to longer-term affective variables (e.g., value of mathematics and expectancy of success) and these latter variables, in turn, are known to predict long-term success in mathematics, e.g., students who value mathematics and have a positive self-concept of their mathematics ability perform better in mathematics classes (Royer and Walles 2007).

While physiological sensors helped to predict student emotion, our recent efforts have moved away from sensors. In part, this movement is because bringing sensors into classrooms does not port to other classrooms, nor scale to many classrooms; e.g., after the sensors are built,

it costs the same amount of time (planning, management) and resources to reach students in each new classroom. Thus we are now experimenting with a different approach, specifically using log data to generate new features of emotion that would allow for sensor-free affect detection, as was also investigated by Baker et al., (2012). Sensor-free affect detection is a more scalable solution, particularly with large numbers of students in public school settings across the country (Baker, R.S.J.d et al. 2012; Wixon et al., in press).

### Animated Affective Learning Companions

The second component in Wayang Outpost that targeted students’ emotion was a suite of animated learning companions that responded to student emotion, see Figs. 1 and 14 (Arroyo et al. 2009b). These full-bodied animated agents acted like peers/study partners who care about student progress, offer supportive advice, and promise to work together with the student, while sitting behind their own desks (Arroyo et al. 2009b. Research



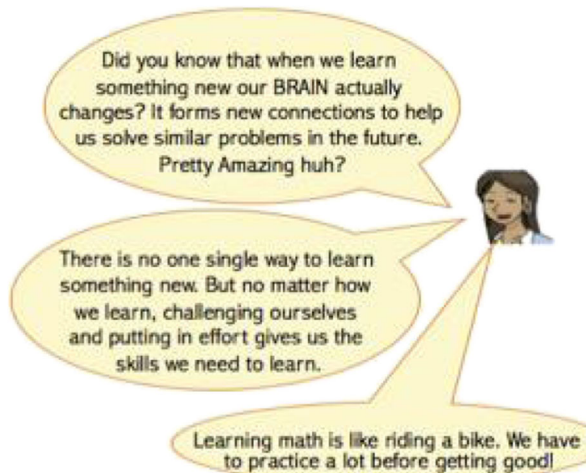
**Fig. 14** Animated pedagogical agents display a range of emotions. Companions act out their emotion and resolve negative ones, expressing full sentences of affective and metacognitive nature, to support growth of mindset towards the view that intelligence is a state (and thus changeable)



questions included: Can human-like learning companions improve motivation and affect? Does the presence of learning companions impact student learning? Are learning companions that resemble a student's gender more effective? How should pedagogical agents respond to affective states or traits of negative valence? Should students be praised when they do well?

We based much of our implementation of the learning companions' dialogue on Dweck's research on human motivation (Dweck 1999, 2002a, b). This theory holds that students who view their intelligence as fixed and immutable (trait-based) tend to shy away from academic challenges, whereas students who believe that intelligence can be increased through effort and persistence (state-based) tend to seek out academic challenges. Students who are praised for their effort are more likely to view intelligence as being malleable, and their self-esteem will remain stable regardless of how hard they have to work to succeed. Thus praise, when delivered appropriately, can encourage students to view their intelligence as malleable and support students' stable self-esteem. However, stakeholders (e.g., teachers, parents) may lead students to accept a trait-based view of intelligence by praising students' intelligence, rather than effort, thus implying that success and failure depend on something beyond the students' control.

Figure 15 presents a few of approximately 50 spoken messages that Wayang's learning companions say to motivate students and also provide metacognitive help. The companions speak the messages either at the beginning of a new problem, or in response to students problem-solving actions. Thus, the companions are non intrusive – they work on their own “computer” (an animated image of a computer) trying to solve the target problem, and react only after the student has entered the problem solution. Some of the messages emphasize perseverance by addressing students' effort in challenging tasks; others debunk myths about the innateness of math ability. Also, companions appear unimpressed or simply ignore students' solutions when students do not exert effort, regardless of success. Exertion of student effort is measured by several variables in the log data, including time to read the problem, number of hints requested, and answer submitted. Companions praise students who exert effort, even if the answers are incorrect.

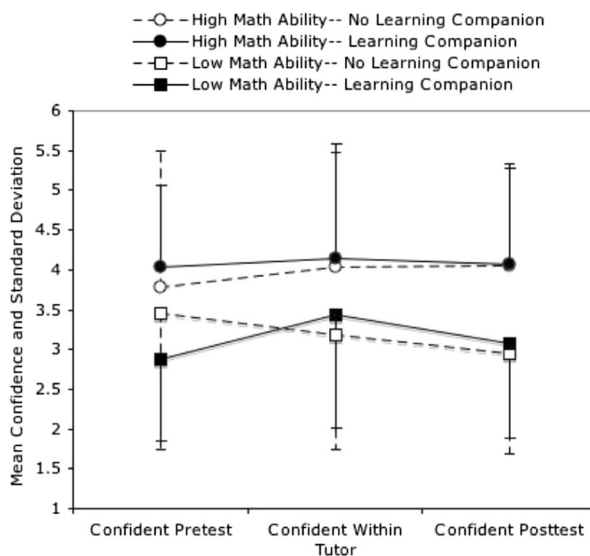


**Fig. 15** Examples of a few of the 50 messages spoken in Wayang Outpost by animated learning companions

Over 100 high school students were assigned to either a learning companion condition (LC group) or to a no learning companion condition (no-LC). The gender of the learning companion was randomly assigned for the LC group (Arroyo et al. 2011b; Woolf et al. 2010). For this study, companions acted upon the tutor's assessment of student effort as indicated in Table 1, and not the student's emotional state. However, emotion self-reports within the tutor were used for gathering additional data on students' emotions as they used Wayang. Both cognitive and affective pre-tests and post-tests were provided to analyze the general impact of learning companions on students' performance and attitudes towards math and learning. The cognitive pre/posttest consisted of math word problems from the MCAS state-wide standardized test in Massachusetts, and the affective pre and posttest consisted of affective predisposition towards problem solving "How [confident/frustrated/excited/interested] do you feel/get when solving math problems?"

In all analyses, which consisted of between-subjects comparisons as part of Analyses of Variance, we accounted for covariates that consisted of the corresponding pretest baseline variables (e.g., we accounted for students' pretest baseline confidence towards problem solving, when analyzing a student's self-report about their confidence within the tutor). Independent variables corresponded to *condition*, specifically learning companion (LC) present vs. absent, or LC type (Female-LC vs. Male-LC vs. no-LC) depending on the analysis. We analyzed both main effects and aptitude-treatment interaction effects for condition and achievement level (math ability, based on math pretest score). In addition, because of the special affective needs of low-achieving students, we repeated the ANCOVAs for the low-achieving student population only, by analyzing a potential "targeted effect" for this group of students alone.

The main impact of learning companions was on affective and motivational outcomes, such as increase of confidence (or reduction of anxiety), see Fig. 16. Students in



**Fig. 16** Students reported their confidence before, during and after using the tutor. Low achieving students (*squares*) showed larger gains in confidence when the companions were available than did high achieving students (*circles*). High achieving students did not change much

general reported more interest (less boredom) when learning companions were present than when they were absent. Significant main effects also indicated that learning companions had a positive impact for all students in general on some measures, e.g., students receiving the female companion in particular had significantly higher math liking and appreciation ( $p < 0.05$ ) and self-concept of their math ability ( $p < 0.05$ ), including expectancy of success, and belief in their current math ability, at posttest time.

Learning companions were especially important for the affective experiences of low achieving students (median-split based on math pretest score). When learning companions were present, low-achieving students reported positive affect nearly twice as often as low-achieving students who did not receive learning companions (for confidence,  $p < 0.01$ , for interest,  $p < 0.1$ ), and it was only when learning companions were absent that a large affective gap existed between low and high achieving students. Similar results that indicate a positive effect of affective characters on low achieving students was found for several outcome variables, including posttest self-concept ( $p < 0.1$ ), perceptions of learning ( $p < 0.01$ ), as well as confidence in their ability to solve math problems ( $p < 0.05$ ). However, learning companions did not help to reduce the gap between high and low achieving students in some respects: compared with high achieving students, low-achieving students still engaged in more quick-guessing ( $p < 0.05$ ) and reported less interest overall while using the tutor ( $p < 0.1$ ), across all conditions.

Students in general improved their math problem-solving performance after working with Wayang (i.e. math posttest score was significantly higher than the pretest,  $p < 0.05$ ), with low-achieving students improving more than high achieving students across all conditions (i.e., posttest – pretest gain comparison between low and high achieving students,  $p < 0.05$ ). Learning companions did not impact student learning directly, but rather induced positive help seeking behavior that had been found to be predictive of student learning in previous studies (Arroyo and Woolf 2005) –specifically, students spent more time on hinted problems ( $p < 0.1$ ), thus either seeking deeper for help, or paying more attention to help, or both. An interpretation of these results is that low achieving students engage in more positive problem solving behaviors *due* to an enhanced affective/motivational impact that encourages focus and perseverance, which is instilled by the affective learning companions. Further details about these results can be found at Woolf, Arroyo et al. (2010a).

### Using Gender Differences to Impact Student Emotion

Another component in Wayang Outpost that targeted student emotion was manipulation of the gender of the learning companion, as it was not clear which kind of character would benefit each student. Gender differences were investigated by focusing on cognitive and affective factors in learning, in relation to whether they were gender matched or unmatched (between student and character gender); students were randomly assigned to male and female companion characters and a no-LC condition.

In general, some of the results showed that characters were positive for both girls and boys, but still some significant results showed that learning companions had impacted only girls positively, but not boys (e.g. and males “quick-guessed” less when characters were absent). Students (both boys and girls) who received the female learning companion reported significantly higher self-concept and liking/appreciation

of mathematics at posttest time as compared to students who received the male character. After more detailed analyses of the data, the effect appeared mostly due to the impact on female students, who scored a full standard deviation lower in frustration within the tutor ( $d=0.99, p<0.001$ ) compared to female students in the control (no-LC) condition. Male students receiving the Jane character, instead, had zero effect regarding frustration when receiving the female character ( $d=0.00$ ) compared to male students in the control (no-LC) condition.

It is important to note that there were no significant differences in mathematics achievement across genders before using the tutor (i.e. math pretest). However, a gender difference was still present for “How *frustrated* do you get when solving math problems?” at pretest time. These high school girls consistently reported lower confidence and higher frustration and anxiety toward solving mathematics problems at pretest time (i.e. affective pretest). Thus, girls in particular especially needed the affective support. In light of this, the results in the previous paragraph regarding the benefit of the female affective character on female students were very welcome as the female character helped to improve the affective reports of girls to a large extent.

In addition, self-reports for the emotion “excitement towards problem solving” at posttest time was higher for female students who worked with companions than for female students who did not receive companions (females in the no-LC condition); in contrast, excitement among male students was higher when companions were absent (males in the no-LC condition). Additionally, girls perceived the entire learning experience with Wayang as significantly better than did boys, in particular when learning companions were present, whereas the opposite was true for boys, who reported better perceptions when the companions were absent. These results suggest that, when the goal is to “reduce frustration” or “increase excitement/interest,” girls should receive the female learning companion and boys should receive the male character or no character at all.

This research into matching the gender of learning companions highlights how to best support female students in intelligent learning environments, but leaves open questions about how to support male students and the reasons for these differences. Perhaps female students should work with female learning companions, male students should receive a male learning companion and high-achieving male students should receive no learning companion at all. This is because of the evidence that low-achieving students (both male and female) benefited from affective learning companions (Woolf et al. 2010). Another possibility is that we should start running focus groups to create new affective digital characters that are especially tailored for boys, from scratch, attempting to understand their expectations of pedagogical agents and avatars.

## Discussion

This article described a landmark learning system, an intelligent adaptive tutor named Wayang Outpost (now *MathSpring*), along with a variety of components used in the system. One important take-home message from this work is that *cognitive*, *affective* and *metacognitive* (CAM) factors can and should be modeled and supported by intelligent tutoring systems. We have shown that a variety of these factors, and

combinations of these factors, influence student behavior within the tutor and student outcomes after using the software. This article also described several evaluations that measured the impact of each component designed to provide a holistic array of supports, based on the cognitive, metacognitive and affective states of the student. This tutor has led to improved performance in mathematics and on state standardized tests, as well as improved engagement and affective outcomes both for groups of students as a whole, and for certain subgroups in particular, e.g., female students and low achieving students.

Given that the Wayang Tutor is traditionally used for short periods of time (i.e., only four or five 50 min sessions on average) the benefits to learning observed throughout the years provides evidence of Wayang's effectiveness, and also argues for its potential use for longer time exposures (i.e., students using Wayang Outpost as a supplement to daily math class).

*Student cognition* was addressed by using several interventions. For instance, a mechanism was created to provide adaptive sequencing of math problems adjusted to students' recent levels of ability and effort exerted (Arroyo et al., 2010b). Despite substantial research on modeling student knowledge, less work has explored how tutors can enforce sequencing of content depending on students' recent performance (with notable exceptions, e.g., Brusilovsky and Vassileva (2003); Karampiperis and Sampson (2005)).

We suggest that our mechanism that adjusts problem selection to both student effort and cognitive expertise is a key contribution to Wayang's success. In addition, the use of math facts retrieval training (MFR) software provided a valuable supplement to Wayang Outpost, for students at all levels. After 3–4 days of tutoring, the MFR software combined with Wayang effectively improved students' performance on items in a mock-standardized test, more than did the tutor alone, and specifically improved performance on the most difficult problems in a math test. Difficult items on these tests generally involved multiple steps and additional computation, and MFR training seems to have freed up memory resources needed to solve those problems. We learned that going back to already mastered pre-requisite topics is an efficient strategy as it can facilitate student learning of more difficult topics later on. We believe that mechanisms such as these, which are based on solid research into human cognition and memory, are important and deserve further exploration by researchers.

*Student affect* was automatically predicted while students used the tutoring system. Initially, this was achieved through information from physiological sensors and student behavior within the tutor. We later created detectors (quantitative models and classifiers) to predict student self-reported emotions. Our recent studies demonstrated detectors that predicted affect without physiological sensors, relying only on behavior patterns and baseline affective traits, which generalize across students in different schools (Shanabrook et al. 2012; Wixon et al. 2014).

A variety of features in Wayang Outpost were designed to help *improve students' affective experience* as they learn with the tutor. For example, affective learning companions trained attributions for success/failure and emphasized the malleability of intelligence and the importance of effort/perseverance. Companions were able to at improve students' affective states (e.g., frustration, confidence) while using Wayang and motivational outcomes after using Wayang (e.g., math liking, expectancy of success in math, self-efficacy), at least when the gender of the character was matched to the gender of the student, and especially for girls and low achieving students. In

addition, our results with the student progress pages suggest that tutoring systems can promote students to transition into affective states of positive valence. Interestingly, this was a *metacognitive intervention* that also helped to resolve and address an *affective problem*.

*Student metacognition* was addressed through activities that encouraged students to reflect on their progress, goals and self-evaluation while in the forethought phase before rushing to the performance phase (Zimmerman and Moylan 2009). This was accomplished by presenting progress charts and tips between problems during the tutoring/practice experience, and also by presenting a student progress page that helps students to reflect on both their mastery and effort, and helps them to make informed choices from there on. The impact of metacognitive supports resulted in improved performance on posttest scores and standardized tests and improved affect and engagement with Wayang.

Using cognitive, affective, and metacognitive teaching strategies is well known in classrooms; teachers, tutors and parents typically consider all of these dimensions when working with K-12 students. In addition, the next generation of intelligent tutoring systems should contain interventions that *act upon* and respond to students' cognitive, metacognitive and affective states as they occur while they work with software, which in turn, will help support cognitive, metacognitive and affective changes within students. An important factor to consider is that the outcomes of specific interventions may impact outcomes across all three areas. For instance, the impact of *metacognitive* pedagogical moves might influence cognitive and affective outcomes as well as metacognitive outcomes. Similarly, *affective* pedagogical moves might influence metacognitive and cognitive student outcomes and behaviors/states. It is very likely that an intervention that is specifically tailored to address one of these dimensions will have a higher impact on that specific area (e.g., affective move on affective outcome), and to a lesser extent to the other two areas (e.g., affective move on cognitive outcome); however, cross-overs are a clear possibility as we have shown in this article.

We believe the next generation of intelligent tutoring systems should reason about the complexities of multiple and sometimes competing goals and outcomes. For instance, it is possible that decreasing the level of challenge of the learning activity (e.g., math problem) might have the benefit of reducing a student's anxiety (affective outcome), but it might not be ideal from a cognitive perspective (as the student might answer easy problems correctly, yet become bored, e.g., work outside their zone of proximal development). We believe that these competing goals and outcomes, and how to prioritize them, present a new set of research questions leading towards a higher level of complexity. As modeling students becomes more complex, new issues become ripe for exploration.

In sum, Wayang Outpost is an important landmark in learning technologies, as it has been designed and evaluated for effectiveness within cognitive, affective and metacognitive dimensions, showing benefits for all of them, and across them. The system presents a rich research environment in addition to a unique adaptive tutoring system that addresses an integral view of human learning, aggregating all three perspectives. Wayang's main pedagogical decision-making method integrates cognitive, metacognitive and affective factors about the student before making a teaching decision based on student learning needs. It provides a real opportunity to contribute to the learning sciences and to our knowledge about human learning.



The main limitation of our approach is the increased complexity of each component design and evaluation. In the future, much more work is needed to conduct research at the intersection of the *metacognitive*, *affective* and *cognitive* perspective, to develop instructional technologies that approach the effectiveness of expert human teachers and tutors, and to provide optimal experiences that instruct, encourage, and generate student agency and promote positive experiences for students while learning new domains about within one learning environment.

**Acknowledgement** This research was supported by the National Science Foundation (NSF) #1324385 IIS/Cyberlearning DIP: Collaborative Research: Impact of Adaptive Interventions on Student Affect, Performance, and Learning and by the Office of Naval Research, # N0001413C0127 STEM Grand Challenges: Building an Effective and Efficient OPEN Tutor Platform. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF or ONR.

## References

- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(3), 277–320.
- Aleven, V., Roll, I., McLaren, B., Ryu, E. J., & Koedinger, K. (2005). An architecture to combine meta-cognitive and cognitive tutoring: Pilot testing the Help Tutor. In *Proceedings of Artificial Intelligence in Education*, 17–24.
- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward Meta-Cognitive Tutoring: A Model of Help-Seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education (IJAIED)*, 16(2), 101–128.
- Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2010). Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist*.
- Arroyo, I., & Woolf, B. (2005). *Inferring learning and attitudes from a Bayesian Network of log file data*. In Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED'05), 33–40.
- Arroyo, I., Beal, C., Murray, T., Walles, R., & Woolf, B. P. (2004). *Web-Based Intelligent Multimedia Tutoring for High Stakes Achievement Tests Intelligent Tutoring Systems*. In Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS'04), 142–169.
- Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Woolf, B. (2007). *Repairing disengagement with non-invasive intervention*. In Proceedings of the 13th International Conference on Artificial Intelligence in Education, 195–202.
- Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009a). *Emotion Sensors Go To School*. In Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED'09), 17–24.
- Arroyo, I., Woolf, B. P., Royer, J. M., & Tai, M. (2009b). *Affective gendered learning companions*. In Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED'09), 41–48.
- Arroyo, I., Cooper, D., Burleson, W., & Woolf, B. P. (2010a). Bayesian networks and linear regression models of students' goals, moods, and emotions. In C. Romero, S. Ventura, M. Pechenizkiy, & R. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 323–338). Boca Raton: CRC Press.
- Arroyo, I., Mehranian, H., & Woolf, B. (2010b). *Effort-based tutoring: An empirical approach to intelligent tutoring*. In Proceedings of the 3rd International Conference on Educational Data Mining, 1–10.
- Arroyo, I., Royer, J. M., & Woolf, B. P. (2011a). Using an Intelligent Tutor and Math Fluency Training to Improve Math Performance. *International Journal of Artificial Intelligence in Education*, 21(2), 135–152.
- Arroyo, I., Woolf, B. P., Cooper, D. G., Burleson, W., & Muldner, K. (2011b, 6–8 July 2011). *The Impact of Animated Pedagogical Agents on Girls' and Boys' Emotions, Attitudes, Behaviors and Learning*. In Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies (ICALT), 506–510.



- Arroyo, I., Muldner, K., Burleson, W., Woolf, B. (2014) Adaptive Interventions to Address Students' Negative Activating and Deactivating Emotions during Learning Activities. *Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies (Volume 2)*. U.S. Army Research Laboratory.
- Azevedo, R., Cromley, J., Winters, F., Moos, D., & Greene, J. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science*, 33, 381–412.
- Azevedo, R., Bouchet, F., Harley, J. M., Feyzi-Behnagh, R., Trevors, G., Duffy, M., Taub, M., Pacampara, N., Agnew, L., Griscom, S., Mudrick, N., Stead, V., & Yang, W. (2011). MetaTutor: An Intelligent Multi-Agent Tutoring System Designed to Detect, Track, Model, and Foster Self-Regulated Learning In *Proceedings of the Fourth Workshop on Self-Regulated Learning in Educational Technologies*. In conjunction with the 15th International Conference on Artificial Intelligence on Education (AIED 2011)
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon.
- Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R., Evenson, S. E., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D.J., Beck, J. (2006) *Adapting to When Students Game an Intelligent Tutoring System*. In Proceedings of 8th International Conference on Intelligent Tutoring Systems, 392–401.
- Baker, R.S.J.d., S.M. Gowda, M. Wixon, J. Kalka, A.Z. Wagner, A. Salvi, V. Aleven, G. Kusbit, J. Ocupaugh, and L. Rossi (2012) Sensor-free automated detection of affect in a Cognitive Tutor for Algebra. In *Proceedings of 5th International Conference on Educational Data Mining*, 126–133.
- Baker, R., Corbett, A., Koedinger, K., & Wagner, A. (2004) *Off-task behavior in the cognitive tutor classroom: when students "game the system"*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04), 383–390.
- Baker, R. S., Corbett, A. T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User-Modeling and User Adapted Interaction*, 18, 287–314.
- Bickmore, T., & Picard, R. W. (2004). Establishing and Maintaining Long-Term Human-Computer Relationships. *Transactions on Computer-Human Interaction*, 12(2), 293–327.
- Brusilovsky, P., & Vassileva, J. (2003) Course sequencing techniques for large-scale web-based education. *International Journal of Continuing Engineering Education and Life long Learning*, 13(1–2), 75–94.
- Bull, S. (2012) *Preferred features of open learner models for university students*. In Proc. of Intelligent Tutoring Systems, 411–421.
- Bull, R., & Scerif, G. (2001). Executive Functioning as a Predictor of Children's Mathematics Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology*, 19(3), 273–293.
- Bull, S., Wasson, B., Johnson, M. D., Petters, D., & Hansen, C. (2012). *Helping teachers effectively support group learning*. In Proceedings of Workshop on Intelligent Support for Learning in Groups, Intelligent Tutoring Systems
- Burleson, W., & Picard, R. W. (2007). Affective Learning Companions. *Educational Technology, Special Issue on Pedagogical Agents*, 47(1), 28–32.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245–281.
- Calvo, R. A., & D'Mello, S. K. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and their Applications. *IEEE Transactions on Affective Computing*, 1(1), 18–37.
- Carraher, D. W., & Schliemann, A. D. (2002). The transfer dilemma. *The Journal of the Learning Sciences*, 11(1), 1–24.
- Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology*, 3, 21–29.
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135(2), 218–261.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, Learning, and Instruction: essays in honor of robert glaser* (pp. 453–494). Hillsdale: Lawrence Erlbaum Associates.
- Conati, C., & Maclaren, H. (2009) Empirically Building and Evaluating a Probabilistic Model of User Affect. *User Modeling and User-Adapted Interaction*
- Cooper, D., Arroyo, I., Woolf, B., Muldner, K., Burleson, W., & Christopherson, R. (2009) *Sensors Model Student Self Concept in the Classroom*. In Proceedings of UMAP 2009, First and Seventeenth International Conference on User Modeling, Adaptation and Personalization, 30–41.
- Cooper, D. G., Muldner, K., Arroyo, I., Woolf, B. P., Burleson, W., & Dolan, R. (2010) *Ranking Feature Sets for Emotion Models used in Classroom Based Intelligent Tutoring Systems*. In Proceedings of the International Conference on User Modeling and Adaptive Presentation (UMAP'10), 135–146.

- Cooper, D., Arroyo, I., & Woolf, B. (2011). Actionable affective processing for automatic tutor interventions. In R. Calvo & S. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 127–140).
- Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User Adapted Interaction*, 4, 253–278.
- Dignath, C., Buettner, G., & Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively?: A meta-analysis on self-regulation training programmes. *Educational Research Review*, 3(2), 101–112.
- D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., et al. (2010). *A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning*. In J. Kay & V. Aleven (Eds.), *Proceedings of 10th International Conference on Intelligent Tutoring Systems* (pp. 245–254). Springer: Berlin/Heidelberg.
- Dweck, C. S. (1999). *Self-Theories: Their role in motivation, personality and development*. Philadelphia: Psychology Press.
- Dweck, C. S. (2002a). Beliefs that make smart people dumb. In R. J. Sternberg (Ed.), *Why smart people do stupid things*. New Haven: Yale University Press.
- Dweck, C. S. (2002b). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways) In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61–87): Academic Press.
- Ekman, P. (1999). *Facial Expressions*. New York: John Wiley & Sons Ltd.
- Eynde, P., de Corte, E., & Verschaffel, L. (2007). *Student Emotions: A key component of self-regulated learning? Emotion in education*. San Diego: Academic.
- Geary, D. C., & Hoard, M. K. (2003). Learning disabilities in basic mathematics: Deficits in memory and cognition. In J. M. Royer (Ed.), *Mathematical Cognition* (p. 2003). Greenwich: Information Age Publishing.
- Gobert, J., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics for science inquiry using educational data mining. *Journal of the Learning Sciences*, 22(4), 521–563.
- Graesser, A., Chipman, P., King, B., McDaniel, B., & D'Mello, S. (2007). Emotions and Learning with AutoTutor. In *Proceedings of Artificial Intelligence and Education*, 569–571.
- Graham, S., & Weiner, B. (1986). Theories and principles of motivation. In D. Berliner & R. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 63–84). NY: Macmillan.
- Greene, J. A., & Azevedo, R. (2007). A Theoretical Review of Winne and Hadwin's Model of Self-Regulated Learning: New Perspectives and Directions. *Review of Educational Research*, 77(3), 334–372.
- Hartfield, E. (1994). *Emotional contagion*. United Kingdom: Cambridge University Press.
- Heffernan, N., Militello, M., Heffernan, C., & Decoteau, M. (2012). Effective and meaningful use of educational technology: three cases from the classroom. In C. Dede & J. Richards (Eds.), *Digital Teaching Platforms* (pp. 88–102).
- Johns, J., & Woolf, B. (2006). A Dynamic Mixture Model to Detect Student Motivation and Proficiency. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, 2–8.
- Karampiperis, P., & Sampson, D. (2005). Adaptive Learning Resources Sequencing in Educational Hypermedia Systems. *Educational Technology and Society*, 8(4), 128–147.
- Kay, J. (2012). AI and Education: Grand Challenges. *Intelligent Systems, IEEE*, 27(5), 66–69.
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education (IJAIED)*, 8, 30–43.
- Lajoie, S. P., & Lesgold, A. (1989). Apprenticeship training in the workplace: Computer coached practice environment as a new form of apprenticeship. *Machine-Mediated Learning*, 3, 7–28.
- Lemaire, P., & Lecacheur, M. (2011). Age-related changes in children's executive functions and strategy selection: a study in computational estimation. *Cognitive Development*, 26, 282–294.
- Lester, J. C., Towns, S. G., & FitzGerald, P. J. (1999). Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents. *International Journal of Artificial Intelligence in Education*, 10(3–4), 278–291.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. *American Psychologist*, 57(9), 705–717.
- Long, Y., & Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education AIED 2013* (pp. 249–258).
- Mathews, M., Mitrovic, A., Lin, B., Holland, J., & Churcher, N. (2012). *Do your eyes give it away? Using eye tracking data to understand students' attitudes towards open student model representations*. In *Proceedings of Intelligent Tutoring Systems*, 422–427.

- Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). *Preschoolers' Precision of the Approximate Number System Predicts Later School Mathematics Performance*.
- McQuiggan, S., Rowe, J., & Lester, J. (2008). *The effects of empathetic virtual characters on presence in narrative-centered learning environments*. In Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI'08), 1511–1520.
- Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's performance. *Journal of Personality and Social Psychology*, 75(1), 33–52.
- Muldner, K., Burleson, W., & VanLehn, K. (2010). "Yes!": Using Tutor and Sensor Data to Predict Moments of Delight during Instructional Activities. In Proceedings of User Modeling, Adaptation, and Personalization, 159–170
- Murray, T., & Arroyo, I. (2002). *Towards measuring and maintaining the zone of proximal development in adaptive instructional systems*. In Proceedings of the 6th International Conference on Intelligent Tutoring Systems, 749–758.
- Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal Human Computer Studies*, 45, 669–678.
- Ogan, A., Aleven, V., Jones, C., & Kim, J. (2011). Pervasive social effects of agent-based instructional dialog. In *In Proceedings of the 15th International Conference on Artificial Intelligence in Education (AIED 2011)* (pp. 238–246).
- Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, 18, 315–341.
- Pekrun, R., Frenzel, A., Goetz, T., & Perry, R. (2007). *The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education*. San Diego: Academic.
- Pekrun, R., Goetz, T., Daniels, L., Stupinsky, R., & Perry, R. (2010). Boredom in Achievement Settings: Exploring Control-Value Antecedents and Performance Outcomes of a Neglected Emotion. *Journal of Educational Psychology*, 102(3), 531–549.
- Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, R., & Strohecker, C. (2004). Affective Learning—A Manifesto. *BT Technical Journal*, 22(4), 253–269.
- Rai, D., Arroyo, I., Stephens, L., Lozano, C., Burleson, W., Woolf, B. P., & Beck, J. (2013). *Repairing deactivating negative emotions with student progress pages*. In Proceedings of the 16th of International Conference on Artificial Intelligence in Education (AIED).
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Robison, J., McQuiggan, S., & Lester, J. (2009). *Evaluating the Consequences of Affective Feedback in Intelligent Tutoring Systems*. In Proceedings of International Conference on Affective Computing & Intelligent Interaction, 37–42.
- Roll, I., Aleven, V., McLaren, B., & Koedinger, K. (2007). Designing for metacognition – applying Cognitive Tutor principles to metacognitive tutoring. *Metacognition and Learning*, 2(2–3), 125–140.
- Roll, I., Aleven, V., and Koedinger, K. R. (2011). Metacognitive practice makes perfect: Improving students' self-assessment skills with an intelligent tutoring system. In Proceedings of the International conference on artificial intelligence in education, 288–295.
- Royer, J. M. (Ed.). (2003). *Mathematical Cognition*. Greenwich: Infoage Publishing.
- Royer, J. M., & Walles, R. (2007). Influences of gender, motivation and socioeconomic status on mathematics performance. In D. B. Berch & M. Mazzocco (Eds.), *Why is math so hard for some children* (pp. 349–368). Baltimore: Paul H. Brookes Publishing Co.
- Shanabrook, D. H., Arroyo, I., Woolf, B. P., & Burleson, W. (2012). *Visualization of Student Activity Patterns within Intelligent Tutoring Systems*. In Proceedings of the 11th International Conference on Intelligent Tutoring Systems (ITS'2012), 46–51.
- Stottler, D., & Domeshek, E. (2005). *Intelligent Tutoring Systems (ITSs): Advanced Learning Technology for Enhancing Warfighter Performance*. In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (IITSEC).
- Tai, M., I., A., & Woolf, B. P. (2013). *Teammate Relationships Improve Help-seeking Behavior in an Intelligent Tutoring System*. In Proceedings of the 16th International Conference on Artificial Intelligence in Education.
- Thomson, D., & Mitrovic, A. (2010). Preliminary evaluation of a negotiable student model in a constraint-based ITS. *Research, Practice in Technology Enhanced Learning*, 5(1), 19–33.
- Tronsky, L. N. (2005). Complex multiplication: Strategy use, the development of automaticity with practice, and working memory involvement. *Memory & Cognition*, 33, 927–940.

- Tronsky, L. N., & Royer, J. M. (2003). Relationships among basic computational automaticity, working memory, and complex mathematical problem solving: What we know and what we need to know. In J. M. Royer (Ed.), *Mathematical cognition*. Greenwich: Information Age Publishing.
- Vygotsky, L. (1978) *Mind in society: The development of higher psychological processes*: Harvard University Press.
- Wang, Y., & Heffernan, N. (2013). *Extending Knowledge Tracing to allow Partial Credit: Using Continuous versus Binary Nodes*. In *Proceedings of the Artificial Intelligence in Education Conference*, 181–188.
- Wigfield, A., & Karpachian, M. (1991). Who Am I and What Can I Do? Children's Self-Concepts and Motivation in Achievement Situations. *Educational Psychologist*, 26(3&4), 233–261.
- Wixon, M., Arroyo, I., Muldner, K., Burleson, W., Lozano, C., & Woolf, B. (2014) *The Opportunities and Limitations of Scaling Up Sensor-Free Affect Detection*. In *Proceedings of Educational Data Mining*.
- Woolf, B. (2009). *Building Intelligent Tutors: Bridging Theory and Practice*. San Francisco: Elsevier Inc./Morgan Kaufman.
- Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, 4(3–4), 129–164.
- Woolf, B., Arroyo, I., Muldner, K., Burleson, W., Cooper, D., Dolan, R., & Christopherson, R. (2010) *The Effect of Motivational Learning Companions on Low Achieving Students and Students with Disabilities*. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS'10)*, 327–337.
- Zimmerman, B. J. (2000). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology*, 25, 82–91.
- Zimmerman, B. J., & Moylan, A. R. (2009) Self-regulation: Where motivation and metacognition intersect *Handbook of Metacognition in Education*.