submitted to AERA 2003
Chicago IL

Learning within the ZPD with the AnimalWatch intelligent tutoring system

Ivon Arroyo, Joseph E. Beck, Carole R. Beal, Beverly P. Woolf
University of Massachusetts, Amherst

1. Objectives

Intelligent tutoring systems offer a way for schools to meet the challenges of providing effective instruction to students with a diverse range of abilities, interests and needs [Self, 1999]. AnimalWatch is a computer-based tutor that uses artificial intelligence techniques to provide individualized math instruction for 9-12 year olds. Math word problems are integrated into narratives about endangered species, to engage student interest and help students appreciate the value of learning math. We present the results of a large-scale evaluation of AnimalWatch with over 350 students that demonstrates how AnimalWatch taught mathematics effectively through the dynamic selection of problems and the provision of help. AnimalWatch was able to select the amount of challenge and difficulty of problems to fit students' zone of proximal development. The emphasis of this conference on accountability reflects an increasing need to demonstrate the effectiveness of educational systems. Finding innovative ways to evaluate educational software is an effort in that direction.

2. Description of AnimalWatch

AnimalWatch is an Intelligent Tutoring System (ITS) for basic arithmetic and fractions, with word problems about endangered species. Thus, it integrates mathematics, narrative and biology. As students work through a series of word problems, AnimalWatch dynamically chooses from a large database of word problem templates, which are instantiated with appropriate operands, depending on the student's current proficiency. AnimalWatch maintains a bayesian-probabilistic overlay *student model* that allows it to make inferences about each student's knowledge as he or she solves problems (such as a very fine-grained report card). Based on these estimations about how students do in relation to the problems that are given to them, AnimalWatch adjusts its problem selection to give students problems that will challenge them, so that they improve with the help provided by the software.

When the student enters an incorrect answer, AnimalWatch provides help (i.e., hints). The first hint provides brief messages, for example, encouraging students to try again or to re-check their work. However, more detailed help is provided if the student continues to make mistakes. AnimalWatch provides interactive hints, which break the problem apart into sub-problems, guiding the student towards a correct answer with the aid of multimedia.

AnimalWatch tutors the following topics: addition, subtraction, multiplication and division of whole numbers; fraction readiness, and addition-subtraction of like and unlike fractions. AnimalWatch contains a "topic network" to represent that certain topics are pre-requisites of others. The network exists so that no hard topic problems (such as fractions) will be shown until easier ones (such as whole numbers) have been mastered. AnimalWatch will give easier problems first and harder problems later, as the student demonstrates that he or she can solve simpler problems with few mistakes. A more extensive description of the AnimalWatch project can be seen at [Beal and Arroyo, 2002].

3. Theoretical framework

Evaluating an adaptive tutoring system is a hard problem mainly because students take different paths depending on their interactions with the system. One approach to determine its educational effectiveness is to verify if the activities provided to the student were within students' zone of proximal development (ZPD). Vygotsky describes the ZPD as "the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or collaboration of more capable peers" [Vygotsky, 1978 pg 86]. In practice, each tutor (human or machine) has limited resources and ability to assist the student, so the "effective ZPD" is defined by the difficulty of tasks possible if the student is given the *available* help. Luckin and du Boulay [1999] call this the "zone of available assistance". The ZPD is neither a property of the learning environment nor of the student; it is a property of the interaction between the two. We say that the student or learning is "in the ZPD" when the student demonstrates efficient and effective learning. Murray and Arroyo [2002] operationalized the concept of ZPD for interactive learning environments: tutoring within the ZPD implies providing activities that are of appropriate difficulty as measured by the mistakes that students make while trying to come through within the system, that should be higher than zero for a student to learn.

We faced the challenge of evaluating the AnimalWatch system to determine its educational benefits for the student. For that purpose, we posed the following questions:

1) were problems hard enough to challenge students' ability but not too hard? (were problems within their ZPD?)
2) were students' mistakes reduced while using the system? (were students able to solve problems more independently than before?)
3) what type and amount of help was more useful?

4. Data sources

AnimalWatch was used by 350 5th and 6th grade students in rural and urban schools in controlled lab experiments carried out in the past 3 years. Students used the system for 2-3 one hour sessions on different days. AnimalWatch logged all of its interactions with the student and the student's responses. Students took a test of cognitive development that demonstrated differences in abilities within each group, providing us with an interesting variety of people. However, the average cognitive development across studies was similar. We integrated the data of these studies to perform a major evaluation of the effectiveness of AnimalWatch.

5. Results

Tutoring within the ZPD through problem selection

AnimalWatch adapted instruction to each student mainly by varying the difficulty of the problems provided. This implied both selecting harder or easier topics for the problems, and also the operands involved in the problems, and the amount of skills required to solve the problems. We claim that being within the ZPD implies that the system should present problems to the student some mistakes will be made [Murray and Arroyo, 2002]. The ideal number of mistakes students are expected to make depends on parameters that should be specified in the system, but a reasonable goal for AnimalWatch would be one mistake per problem.

If the average number of mistakes that a student made on the nth problem seen is plotted, the result should be a set of points scattered around a mean of the ideal number of mistakes. The results can be seen in Figure 1. This figure shows that students consistently make about 0.75 mistakes all along the 90 problems (average problems seen per student). During the first 15 problems, however, the system underestimates how much the student knows until it catches up with the student's knowledge, so that the average mistakes increase from 0.5 to 0.75[1]. The 0.75 +- 0.2 average seems a reasonable behavior for the system (a non-adaptive system would present random points all over the graph). We conclude that AnimalWatch was selecting problems within the students' ZPD, as problems were not too hard and not too easy.

---

[1] Even though this behavior could be improved with a better student model initialization, there will always be a period where the system catches up with the student's knowledge.

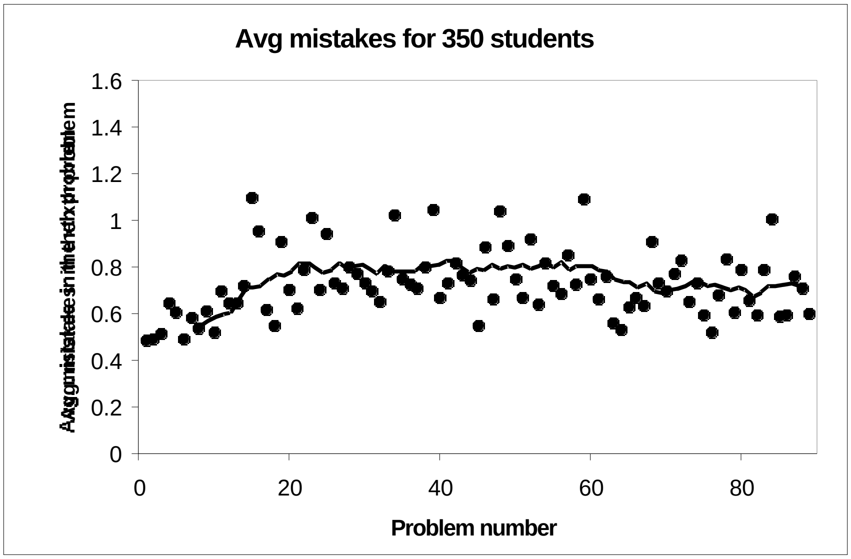


Figure 1. Average mistakes in the nth problem seen across 350 students. (the line is a running average with a period of 10)

AnimalWatch's problem selection mechanism seems appropriate for students' learning. However, this analysis is still incomplete to appreciate actual learning. If students did not improve their mistakes in similar subsequent problems, then selecting appropriate problems is not enough for effective tutoring. The next section performs a deep analysis of students improvement.

Effectiveness of AnimalWatch at various topics

In order to analyze students' learning, we observed students' mistake change at subsequent problems for each topic. This is a hard task as students received very different problems. Our approach was to observe how students' mistakes changed over time at instances of similar problems. We selected only a subset of the 350 students for this analysis, those who had made mistakes in the first problem of each topic, and followed their mistake behavior at the 2^{nd}, …, 6^{th} problems for each topic and difficulty (some topics are divided in *easy* and *hard* to make sure that the problems we consider in each sequence are not too different). For instance, the topic called "easy division" will contain one sequence of 6 problems per student that involved division of at most 3 digit numbers by one digit numbers. This provides some level of control over the difficulty of problems, so that the average difficulty of the 1^{st} through 6^{th} problems are similar (average operand size is not significantly different). This allowed us to observe whether students' mistakes declined after interacting with the help in the system.

Figure 2 shows that students' mistakes did reduce over problems. The overall mistakes after six problems is about 2/3 of the initial mistakes. The graph shows average mistakes at the nth problem over 200 students in the whole number problems and averages of about 100 students in the fraction problems. We attribute this improvement to a combination of the help they received and the problem selection mechanism.

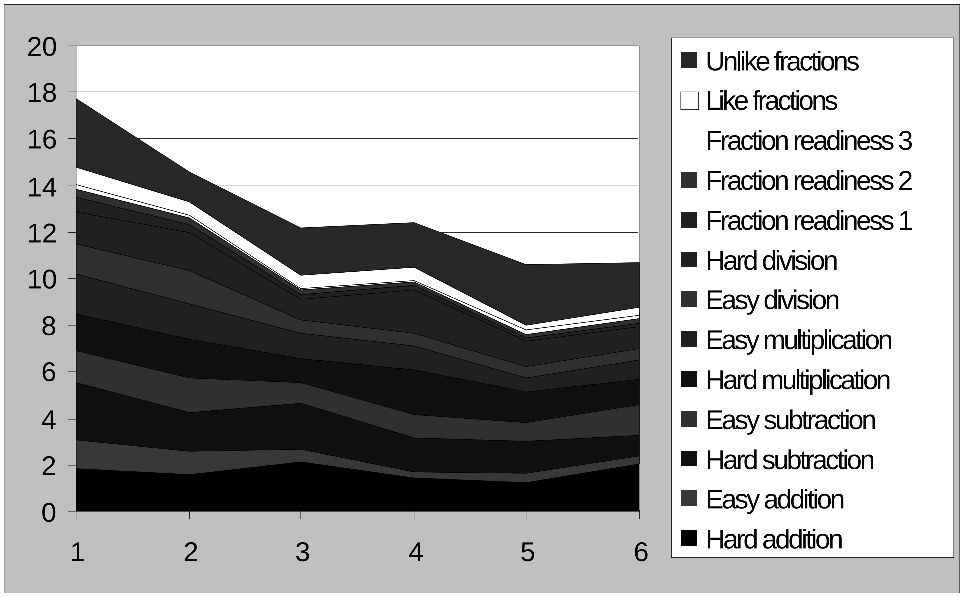
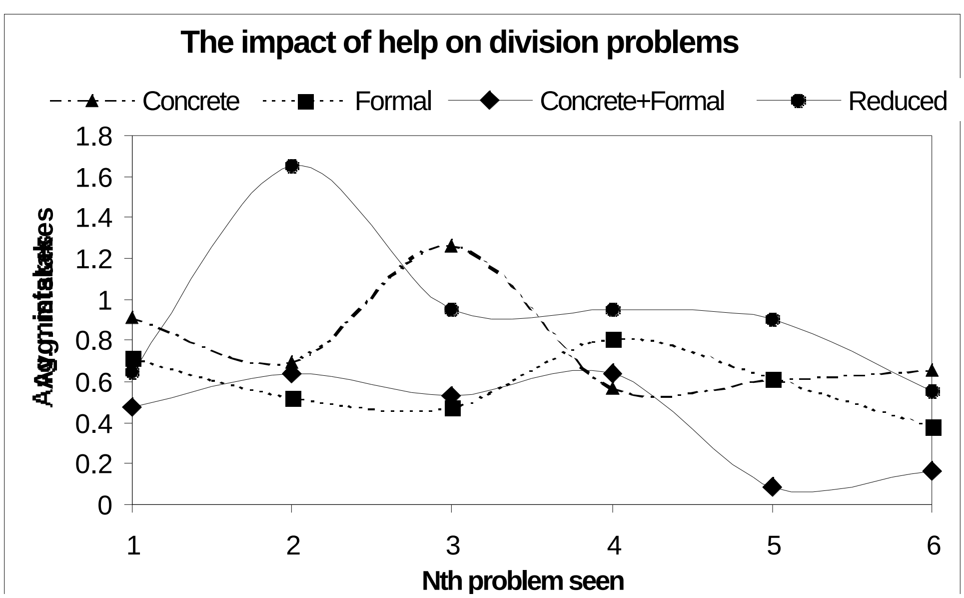


Figure 2. Accumulated mistake reduction in the first six problems seen for each topic

The effectiveness of the interactive help

The hints provided in Animalwatch could be classified in three different kinds: formal-numeric help, concrete-conceptual help, and text messages. The 350 students in the previous sections used four slightly different versions of AnimalWatch. A reduced-help version provided short messages with little support (e.g. *"are you sure you are adding 1/2 + 1/3?")*. The formal-numeric version



provided help in the form of numeric procedures, while the concrete version used cuisenaire-rods to illustrate a concrete-conceptual solution. Finally, the Concrete+Formal version of AnimalWatch gave both concrete and formal-numeric help before giving out the answer. We wanted to analyze whether the different kinds of help were effective and what amount of help was better.

We thus picked one topic in the system: whole number division. We chose this topic because it has medium difficulty, and thus both high and low achieving students could have made mistakes on it (this is probably not the case for addition).

The results in figure 3 are average mistakes at the nth easy division problems (division by one digit). We only considered students who had seen six division problems of this kind. This restriction resulted in 100 students with their mistake behavior at 6 easy division problems. An overall mistake reduction trend can be observed in figure 3. It can also be observed that students in the Concrete+Formal version reached mistakes close to zero after the fourth problem.

When considering an average of the overall mistakes made by each group in the six problems, students in the Concrete+Formal version made the least mistakes and students in the Reduced help version made the most mistakes. More importantly, an average mistake reduction was computed per student by averaging the differences of mistakes between each problem and the following one. Then, an average mistake reduction was computed for each group, which is also shown in table 1. Mistake reduction was significantly worse for the reduced-help group while it was significantly better for the Concrete+Formal group.

Table 1. Average mistakes and average change in mistakes for the different help types

	N	Avg. mistakes	Avg. Change
Concrete	23	0.68	-0.13
Formal	21	0.51	-0.37
Concrete+Formal	36	0.32**	-0.52**
Reduced	20	0.81*	0.20**
Total	100	0.54	-0.26

* t-test, significantly different that the other groups, $p < 0.05$
** t-test, significantly different that the other groups, $p < 0.001$

We conclude that the help provided in AnimalWatch was effective at helping students learn mathematics. The version of AnimalWatch that made students improve the least was the one providing reduced-help. Meanwhile, the more informative and interactive help was most effective: the best AnimalWatch was the one providing the most help, combining formal-numeric and concrete-conceptual help. Similar results have been seen for other topics.

6. Conclusions

AnimalWatch has shown to be an effective piece of software for teaching basic mathematics. Mistakes get reduced overall, specially when more help is provided (a combination of concrete and formal help supplementing each other). Mistakes get reduced the least when less interactive help is provided. We have also shown that AnimalWatch maintains a constant level of challenge for students, even though they are improving while using our tutor. This research provides evidence for the effectiveness of Intelligent Tutoring Systems. It is also a contribution to the measurement of gains in systems that are extremely dynamic, after comparing the behavior of subjects that went through very different paths but at some point faced similar situations. Moreover, AnimalWatch has been well accepted by students and teachers, and it has proved to improve students' attitudes towards mathematics [Beck et al, 1999].

7. References

Beal, C. R., & Arroyo, I. (2002). *The AnimalWatch project: Creating an intelligent computer mathematics tutor.* In S. Calvert, A. Jordan, & R. Cocking (Eds.), Children in the digital age. Greenwood.

Beck, J. E.; Arroyo, I.; Woolf, B. P.; Beal, C. R. (1999) *An Ablative Evaluation*. In Proceedings of the Ninth International Conference on Artificial Intelligence in Education. pp. 611-613. IOS Press.

Luckin, R; du Boulay, B. (1999) *Ecolab: the development and evaluation of a Vygotskian design framework*. International Journal of Artificial Intelligence in Education, 10, 1999, pp. 198-220

Murray, T., & Arroyo, I. (2002). *Towards Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems.* In Proceedings of the Sixth International Conference on Intelligent Tutoring Systems. Lecture Notes in Computer Science. Springer.

Self, J. (1999). *The Defining Characteristics of Intelligent Tutoring Systems Research: ITSs Care, Precisely*. International Journal of Artificial Intelligence in Education, 10, 1999, pp. 350-364.

Vygotsky, L.S. (1978). *Mind in Society: The development of higher psychology processes.* Cambridge MA: Harvard University press.