Animalwatch: an arithmetic ITS for elementary and middle school students

Ivon Arroyo[1] - ivon@cs.umass.edu

Department of Computer Science - School of Education

Computer Science Building - University of Massachusetts, Amherst, MA 01003-4610

1. Introduction

Animalwatch is an Intelligent Tutoring System (ITS) that was developed with two objectives in mind: 1) to effectively teach pre-algebra arithmetic to elementary and middle school students, 2) to increase young girl's confidence, value, and liking of mathematics in the process (Beal, 2000). Animalwatch integrates mathematics with science, language arts, and technology to achieve its goals. In several experimental trials with 9 to 11 year-old students, Animalwatch increased students' liking, confidence and value of mathematics (Beck, 1999). The purpose of this paper is to report on Animalwatch as a mathematics learning tool and on how we intend to improve its original design to address users' learning needs and expectations about its functioning. Moreover, I intend to show at each step how our ideas can also be applied to the algebra tutoring systems domain.

2. Animalwatch's interface and internal structure

Our ITS's implementation details will be discussed in this section, and also how it feels for a student-user to interact with Animalwatch.

2.1 Interacting with Animalwatch

A class of students arrives at the computer lab to run Animalwatch. They find a first screen where they choose an endangered animal species. After that, students read a letter from the World Wildlife Foundation on the screen that assigns them the role of a scientist who is in

[1] *The rest of the Animalwatch crew: Joseph E. Beck, David Marshall, Carole R. Beal, Beverly P. Woolf*

charge of performing research about the species. From there on, students work on math word problems about the chosen animal. Students use mathematics to solve environmental and ecological problems as they observe, monitor, and manage the endangered animals (see Figure 1).

Whenever students enter a wrong answer to the problem, the system gives them immediate feedback with increasing levels of information. The first hints will be short messages encouraging the student to try again. However, if the student keeps on making mistakes, Animalwatch will eventually provide hints that guide the student through each of the multiple steps involved in solving the problem (see Table 2 for examples on hints).

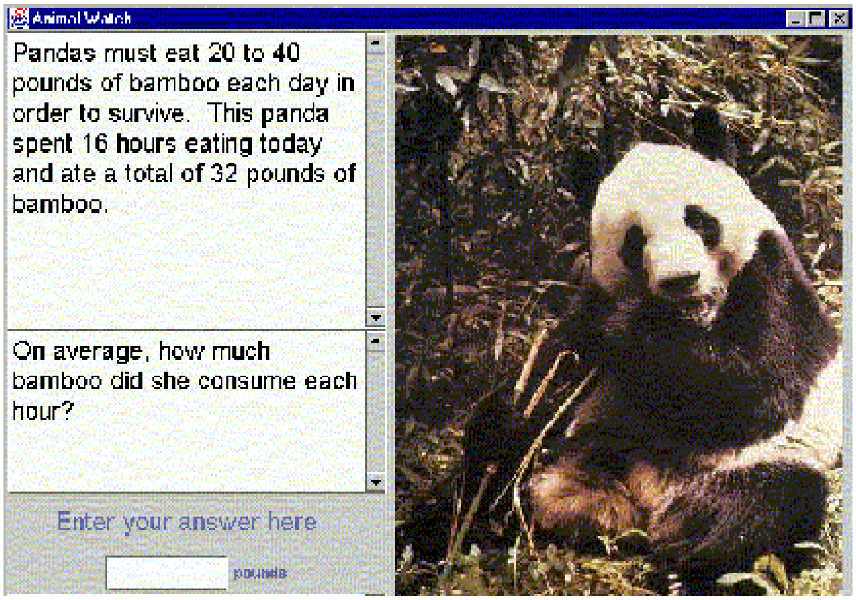


Figure 1: A whole number division word problem in Animalwatch

2.2 Animalwatch's internal structure

Principles of construction. Animalwatch runs both on PC and MAC platforms to address school equipment requirements. Students can use it in multiple sessions as their student models will be saved from session to session. If it is not the first time that the student logs in, Animalwatch will take the student to the point where he left off in the prior session. It can save student models to a server in the school computer lab, so that the following day students can sit on any computer to continue working where they left off. Animalwatch was developed in Java, and works as a standalone application. In our dissemination phase, it will be available for download from a web page.

Student model and domain knowledge. Animalwatch tutors addition, subtraction, multiplication and division of whole numbers, introduction to fraction concepts, addition of fractions, and subtraction of fractions. It maintains a student's proficiency estimate for each topic in the domain. Animalwatch is intelligent in that it provides students with problems for which the student is "ready". For example, a student is ready to see a word problem that involves fractions only after she has shown proficiency in all the topics that involve whole numbers. Also, there is a notion of difficulty of a problem within a topic that depends on other factors (size of the operators, number of steps involved in solving the problem, etc.). More details about the student model can be found in (Beck, 1997). Animalwatch has been designed so that it is easy to add in new topics into the system. That is why it would be easy to extend it to algebra. It would require very little programming to extend the ITS to include problems of the type $2x + 3 = 11$

3. The future: Animalwatch's improvements

Although our ITS was positively accepted in the school experiments, we found new requirements both from students and teachers that we did not expect. This section introduces these needs and explains how we intend to address them. It also introduces our research on how to make Animalwatch a more effective teaching tool.

3.1 Story line: keeping students interested with a goal.

Our goal is not only to have students learn mathematics, but also to raise their interest in it. Prior versions of Animalwatch gave students one problem after another together with some visual reward when problems were correctly solved. However, after several trials in schools I encountered comments as "the system doesn't have a goal", or "the program has no end". I consequently decided to build a story line into the system, so that the concrete goal for students would be to reach the end of the story. The interaction of the student with the new system goes as follows: students choose an animal at the beginning of the session, to then sequentially go through various "story contexts" that last for a fixed number of minutes, until they get to the end of the story. For example, students choosing the Giant Panda will go through the following contexts: first, they will go to the Washington Zoo to learn about Pandas in captivity, then they will raise money for a trip to China and pack, and last they will go to a Panda reserve in China to study the Pandas in the wild. Within each section, students are given problems which are intelligently chosen depending on students' readiness for a topic and a difficulty level within that topic. This way, two students that are in the same point of the story at a particular time will be covering different mathematics material. While one student may be working on whole number multiplication, the other may well be working on addition of fractions.

Keeping students' interest is one of our major challenges. First, because one of our research goals is to achieve girls' and boys' interest in math with our software. Second, because we think that interest in using our software also will affect learning. Learning math did not seem to be an appealing enough goal by itself, so we had to build in a more attractive goal for them o pursue, given students' age and interests. I think algebra educational software also needs to address the issue of keeping students' interest, and adding a story line in this way is one possibility.

3.2 Problem repetition and the word problem database.

Animalwatch has a large database of templates for word problems. These templates can be displayed with different numbers to make the problem harder or easier to solve. A difficulty we encountered was that students complained about getting one same problem twice but with different numbers, so we had to guarantee that problems were not seen more than once. Given this constraint, our word problem database became larger. In addition, because problems for each one of the story-contexts and for each animal and for each topic are needed, a combinatorial explosion of word problem generation is inevitable. We currently have 600 word problem templates plus images that go with them, but we are planning to grow it to 800. I have started to consider the implementation of a word-problem authoring tool. Moreover, teachers who will be Animalwatch users are an invaluable source of word-problem generation expertise, so they should be able to incorporate new problems and share them with the community of Animalwatch users in different schools. I am still thinking about how this tool should be implemented to make it easy for teachers to use. I wonder how other word-problem based ITS have addressed this tedious task of word problem generation in ITS, as I don't know of any published papers on this topic.

3.3 Hint effectiveness research

We consider that a hint is effective if the student makes fewer mistakes after being exposed to it. The following section describes some results on hint effectiveness in Animalwatch, and new ideas on how to improve immediate feedback.

3.3.1 Analyzing hint effectiveness

These are results for 70 students who used Animalwatch for 3 hours in a rural area school of Massachusetts. Hint effectiveness was measured by analyzing how the number of mistakes made changed after seeing a hint. I considered N=472 cases where the student saw a problem, made mistakes, saw some hints, and later got a new similar problem[2]. I analyzed how the number of mistakes changed from the first problem to the following one (see Table 1). The number of mistakes made from one problem to the following one decreased, and the difference was significant (paired samples t-test, $p<0.000$). I conclude that hints were effective on average, as students tended to make fewer mistakes after seeing the hints.

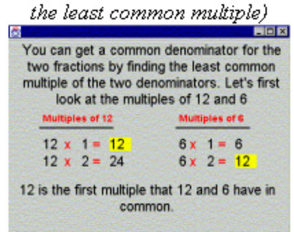
	Wrongly answered problem	Following similar problem
Average number of mistakes	2.25	0.61
Average problem difficulty level	1.51	1.51

Table 1. Means for number of mistakes and difficulty levels in pairs of similar contiguous problems

3.3.2 Adapting hint selection for groups of students

I think that one way to make our system more effective at teaching is to design different types of hints, i.e. to find groups of students that behave differently depending on what hint they obtain, and then to include that reasoning into the system at the moment of hint selection.

Having this in mind, I designed more than one hint to aid a particular skill. These alternative hints differ from one another on: the amount of information given, the degree of interactivity, and the symbolism level (see Table 2).

	CONCRETE HINTS	SYMBOLIC HINTS
INTERACTIVE	Message hints that talk about "things": *Divide all the things that you have into 5 groups. How many things are there in each of these groups?* *Are you sure you are adding 3 things plus 8 things?*	*Hints that provide highly numeric explanations (finding the common denominator of a fraction by looking for the least common multiple)* 

[2] Two problems ... el within that topic.

For example, 2 ... ers.

I again measured hint effectiveness by analyzing how the number of mistakes changed after seeing a particular hint. I found interaction effects between cognitive development and symbolism in predicting hint effectiveness. These effects suggest that low symbolic hints are more effective for low cognitive developed students than highly symbolic ones. They also suggest that highly symbolic hints are more effective than low symbolic ones for high cognitive developed students. This was a reasonable result as students at lower stages of cognitive development are known to have a hard time with highly symbolic manipulations (Piaget, 1964). This allowed to validate our model in the first place, and also to show how developmental psychology research can be incorporated into an ITS's decisions.

We also found interaction effects between gender and hint interactivity in predicting hint effectiveness. They suggest boys did better with our low interactive/low intrusive hints, and that girls did better with our highly interactive hints. The experiment and the results are explored further in a paper in the conference proceedings (Arroyo, 2000).

I want to modify Animalwatch to incorporate these findings, by providing different degrees of symbolism and interactivity depending on cognitive development level and gender. Cognitive development can be diagnosed with an automated pre-test (Arroyo, 1999), and we can ask students their gender at the beginning of the first session. I intend to perform a new controlled evaluation study. The new hypothesis will be that when these hint selection rules are incorporated in Animalwatch, its effectiveness should be higher for all students.

Extending this research to algebra education shouldn't be hard. We are talking about the same domain and similar ITSs. For example, PAT also works with word problems and provides immediate feedback as a remediation mechanism (Koedinger, 1997). Also, students who are learning algebra are likely to handle different levels of abstractions, especially taking into account that Piaget's formal stage of cognitive development is known to be reached within a broad range of ages (from 11 years old to adulthood) (Case, 1992).

Given our empirical findings, I think that there could be other important qualitative dimensions in hint design that could affect students' learning. I am considering the possibility of performing factorial analysis on hint features (number of textboxes, graphics, number of interactions required, etc.) to look for new relevant dimensions along which to classify hints. I believe this is an entire area of research that has yet to be explored.

3.3.3 Hints to aid on problem solving

Even though our hints have shown to be useful, I still feel they could be better. Problem solving involves at least three steps: 1) Finding the relevant information in the word problem (filtering out irrelevant information, finding out what numbers are involved), 2) deciding what operation to perform with the operands, 3) Executing the operation correctly. I are currently providing hints only for the last problem solving step in Animalwatch.

I are looking for ways to help students with these two first steps. Because short hints like "try again!" have shown to be very effective as first hints (the 45% of correct answers after a mistake was made are thanks to the intervention of this kind of hint!), I want to keep on showing them as a first hint. Thus, I plan on giving a second hint where students would be asked to enter the equation in text boxes. If students fail at this step, I know that I should remediate on the first problem solving step. If students succeed at this step, then I can give them hints to remediate in the third problem solving step.

Still, teaching students how to find the relevant information in the text and what operation to do with them is not an easy task, given the amount of word problems that we have. Solutions to these word problems are very dependent on the text, so it is hard to find a generic way to help students in building equations out of a word problem. I am now thinking of underlining the relevant part of the text to help students filter out irrelevant information, as one possible solution.

3.4 Teacher needs

Teachers have been willing to have our product. However, they also have new demands for us. Teachers would like to monitor students' achievement via on and off-line reports, and maybe accompany the system with personalized and printable exercise sheets to give to students that they can correct by hand (teachers still need to have the feeling they are in control!). Teachers want to be able to customize the system so that it starts at a specific topic, even though it then can go back to more basic topics if the student needs it. There is a growing need for a tool for the teacher to accompany and administer the ITS.

4. Conclusions

This paper has addressed our current work and future directions with Animalwatch, an arithmetic intelligent tutoring system aimed for elementary and middle school students. I have presented some results on hint effectiveness in our system, and also proposed some ways to achieve more effective tutoring with the use of a mapping of type of student-kind of hint, that implies detecting qualitative different students and designing qualitatively different hints. I have also put forth both students' and teachers' new demands to our ITS, and also possible ways to address those needs. Students' needs have to do with affective-emotional factors, essentially keeping their interest in using the system. Teacher demands have to do with making the ITS become not only a tool that will help students learn, but also a student evaluation tool for the teachers themselves. Many of these concerns, difficulties and possible solutions are applicable to algebra learning systems, given various similarities between Animalwatch and well-known algebra tutors.

Arroyo, I.; Beck, J.; Beal, C.; Woolf, B.; Schultz, K.: Macroadapting Animalwatch to gender and cognitive differences with respect to hint interactivity and symbolism. In the *Proceedings of the Fifth International Conference on Intelligent Tutoring Systems.* 2000.

Arroyo, I.; Beck, J.; Schultz, K.; Woolf, B.: Piagetian Psychology in Intelligent Tutoring Systems. In the *Proceedings of the Ninth International Conference on Artificial Intelligence in Education.* pp. 600-602. 1999.

Beal, C. R, Woolf, B. P., Beck, J. E., Arroyo, I., Schultz, K., Hart, D. M.: Gaining Confidence in Mathematics: Instructional Technology for Girls. *In the of Proceedings of International Conference on Mathematics/Science Education and Technology. pp. 57-64. 2000.*

Beck, J. E., Stern, M. J, Woolf, B. P.: Using the Student Model to Control Problem Difficulty.*In the Proceedings of Sixth International Conference on User Modeling. pp. 277-289. 1997*.

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

Case, R.. *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale NJ: Erlbaum. 1992.

Koedinger, K.R., Anderson, J. R., Hadley, W. H., Mark, M. A. Intelligent Tutoring goes to school in the big city. *In Proceedings of the Eighth International Conference on Artificial Intelligence in Education.* pp. 30-43. 1997.

Piaget, J. *The Child's Conception of Number*. Routledge & Kegan. 1964