Piagetian Psychology in Intelligent Tutoring Systems

Ivon Arroyo, Joseph E. Beck, Klaus Schultz, Beverly Park Woolf
{ivon,beck,schultz,bev}@cs.umass.edu
School of Education
Computer Science Department
University of Massachusetts, Amherst

In this paper, we describe the use of Piaget's notion of cognitive development in the building of pre-tests that would allow improving a tutor's reasoning ability. We are interested in finding individual differences that not only predict a student's overall performance, but that can also be easily applied to actual tutoring decisions. Our hypothesis was that students with different levels of cognitive development should behave differently in the tutor, and that this is the reason why they need to be taught with different strategies. We thought it was very likely that our population of elementary school students would have different cognitive levels, so that this feature would be an essential aspect to take into account to adapt the response of a tutoring system. We have adapted classic Piagetian tasks used to measure level of cognitive development for use on computer. We found that this measure predicts student performance at a variety of grain sizes: understanding of hints, amount of time to solve problems, failure rate and also the number of problems students need to attempt to master a topic. We also describe how these measures of cognitive development can be usefully applied to improve the behavior of the tutor for students at different cognitive levels.

Keywords: Cognitive resources, Piaget, pedagogical decisions, individual differences.

Ivon Arroyo
Ivon@cs.umass.edu
Computer Science Department, University of Massachusetts, Amherst, MA 01003
Telephone: (413) 545-0582
Fax: (413) 545-1249

1.Introduction

When we are were able to predict individual differences with computer-based pre-tests, flexibility and adaptability can be improved in a tutoring system, that will certainly benefit students. This is the reason why it is essential to be able to record individual differences in intelligent tutoring systems. We need to build appropriate computer-based pre-tests that are predictors of individual differences to obtain such a broad classification. This will enable to understand factors that influence user's understanding and behavior in the system. Our hypothesis was that students with different cognitive levels should behave differently within the tutoring session. In this paper we describe our design and implementation of pretests for an intelligent tutoring system that would allow to predict cognitive development level (Piaget, 1953). We also examine the results collected after giving the pre-tests to 60 sixth grade elementary school students. We intend to find information that predicts a student's overall performance to be applied in various pedagogical decisions and in the student model.

2. The domain and the experiments

MFD (Mixed numbers, fractions and decimals) is an intelligent tutoring system (ITS) aimed at teaching fractions, decimals and whole numbers to elementary school students (Beck et. al, 1997). A version of MFD was evaluated in May 1998. It tutored operations with whole numbers and fractions. This version was tested with 60[1] sixth grade elementary school students during three days (for a total of three hours using the system). Students were randomly divided into an experimental and a control group. The experimental group used a version with intelligent hint selection and problem selection. Intelligent problem selection consisted in giving the student a problem with an appropriate difficulty level, depending on the level of mastery of different skills. Intelligent hint selection consisted in determining the most appropriate amount of information to provide in a hint. The control group also used a version with intelligent problem selection but received no feedback other than a prompt to try again after an incorrect response. An objective of the current study was to see what benefits (if any) the intelligent help system was providing. In addition, we wanted to investigate the benefits of the intelligent help component when the student was at a particular cognitive level.

We gave the students a computer-based pre-test that measured their level of cognitive development. Ten computer-based Piagetian tasks measured different cognitive abilities. These tasks were selected among various one. Some of them were easier to implement on computer than other ones[2]. We intended to determine with these tasks if the students were at one of the last two stages of cognitive development proposed by Piaget (concrete operational stage and formal operational stage). [3] Seven tasks were given to the students to verify dominance of concrete operations and three tasks checked for formal operations. All these experiments are based on those that Piaget used (Piaget, 1953, 1964; Voyat, 1982; Ginsburg et. al, 1988). The tasks involved a high level of interactivity and they were implemented in Java 1.1.0. Figure 1 shows some screenshots of these tasks. They tested:

- *Number conservation:* Students initially observed two identical sets of cookies (each set consisted of nine cookies horizontally aligned). When the elements of one set were moved to form a small circle, students were asked to determine if the amount of cookies in this last group had changed.
- *Serialization*: Students had to order a group of six pencils from the shortest to the longest one.
- *Reciprocity:* Students were initially presented with two identical vessels with the same amount of liquid. Each of these containers had another empty one next to it: one was very narrow and the other one was very wide. We

[1] Due to absentees among students, we only have complete data for 46 students.

[2] One of the authors (Klaus Schultz) is an expert in Piagetian psychology applied to education, and was very active in the many decisions that needed to be taken with respect to the tasks. These decisions had to do with deciding what tasks to use, and how to modify them to be implemented in the computer, without losing the essence of what the tasks were evaluating. This was not an easy task.

[3] We are aware that administering tasks in this format does not provide the richness of information on students' cognitive development that would be possible with individual clinical interviews. In particular, we have not obtained any information concerning students' reasons for the responses they give –which, in the Piagetian framework, are at least as important as the responses themselves–. Since in this case the determination of students' level of cognitive development is not an end, but a means to the end of more effective tutoring, we believe the approach is justified.

asked students to click on the empty vessels in the approximate place where they thought the level of water was going to be if the liquid from each of the two identical vessels was poured into them.

- *Area conservation:* Students were asked to compare two areas of the same size but different shape.
- *Class inclusion*: Students had to determine whether there were more dogs or more animals in a set with different kinds of animals, in which the largest subset was dogs. After that, and even if the response was correct, they were asked if the response was sustained even if more dogs were added behind a wall that was shown (they couldn't determine how many dogs were going to be added, but there was always going to be more animals than dogs).
- *Functionality*: Students had to invent an algorithm to solve a problem of ordering six pencils by length when they could only see two of them at a time.
- *Reversibility:* Students were shown an animation of three colored balls entering inside a can from one end, one after the other. After that, they were asked to determine the order in which the elements would come out of the same end of the can.

Three more tasks were administered to determine whether the child was at the formal operations stage. We measured:

- *Establishment of hypotheses, control of variables in experimental design, drawing of conclusions:* These were measured with a simulation of plant growth experiments under various conditions of temperature and illumination. The children were asked if it was better to water a plant once a week or once a day. They had also two options of temperature to give to the plant (60 and 90 degrees). They could experiment with four plants by choosing either temperature value and watering frequency. After that, students were asked what their conclusion was.
- *Proportionality*: Students were shown two animals of different heights and were given two different measurement system units (large buttons and small buttons). Students were asked to measure one of the animals with the two measurement units and the other animal with only one of the measurement units. Then, they were asked to infer the height of the last animal with the second measurement system.
- *Combinatorial analysis*: Students were asked to generate combinations of four switches to open a safe. We evaluated if the student could build the sixteen combinations with those switches. The safe opened when the student succeeded at doing this or after a maximum number of trials.

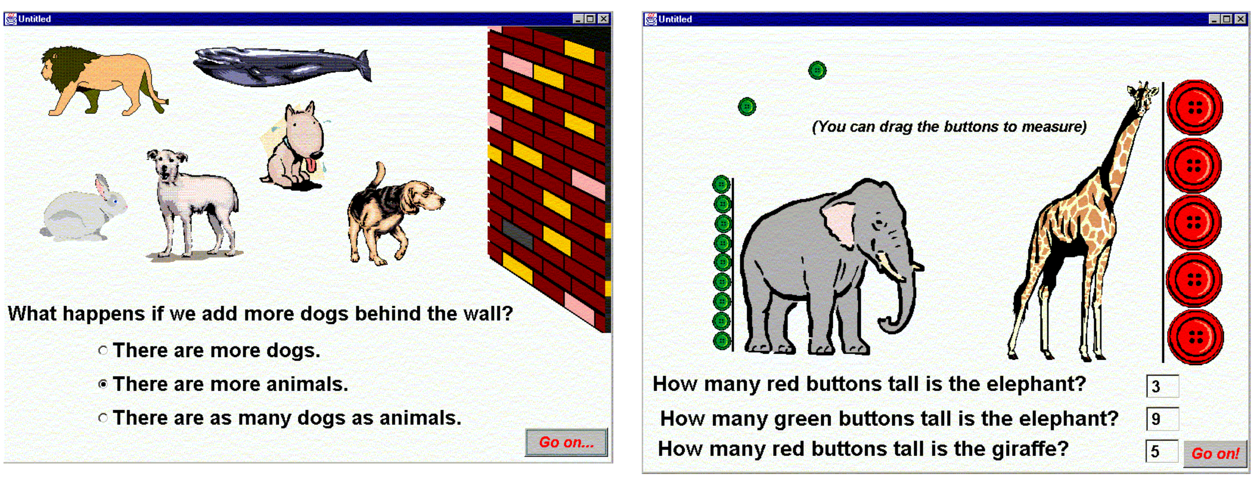

Figure 1: Screenshots of the tasks that tested class inclusion and proportionality

We gave an order to the tasks from what we considered easiest to most difficult. The tasks described above were listed in this order, from easiest to most difficult. Our criterion for this order followed the lines of (Piaget, 1953) but also by our common sense in some cases. Not all the tasks were administered to the student. Every child started the pretest with a middle-level task (reversibility) and were moved to easier tasks when they failed or to more difficult tasks when they succeeded. This way, students with high cognitive levels didn't have to go through tasks that we were almost sure that they would have succeeded at. In addition, students with low cognitive levels didn't have to go through tasks that we were almost sure that they would have failed at. If the student had succeeded at two contiguous tasks that were more difficult than another one we had not shown yet, then we could infer he would also succeed at that one. In addition, if the student had failed at two contiguous tasks that were easier than another one we had not shown yet, then we could infer he would also succeed at that one. Table 1 and table 2 show an example of this method for some student's responses.

task 1	task 2	task 3	task 4	task 5	Task 6	task 7	task 8	task 9	task 10
-	-	succeeded	succeeded	failed	succeeded	failed	failed	-	-

Table 1: This representation of failure and success in different tasks for some student shows that tasks 1, 2, 9 and 10 will not be administered.

task 1	task 2	task 3	task 4	task 5	Task 6	Task 7	task 8	task 9	task 10
succeeded	succeeded	succeeded	succeeded	Failed	succeeded	failed	failed	failed	failed

Table 2: Inferred results for the shown and not shown tasks

3. Description of results

The number of Piagetian tasks that the student accomplished was used as a measure of cognitive development, being the minimum possible level equal to zero and the highest possible level equal to 10. The mean number of correct answers for the sixth grade pupils in the study was 5.7 (min. value=0, max. value=9), with a standard deviation of 2.1. Most students could do approximately half of the tasks correctly, which makes us believe that our estimation of the approximate number of tasks that they succeed in was appropriate. The mean number of correct responses is independent from sex and condition: there is no significant difference between the tasks boys and girls accomplished (girls' mean correct responses = 5.2; boys' mean correct responses = 5.9), or between control and experimental groups (control group's mean correct responses = 5.7, experimental group's mean correct responses = 5.5).

4. Is cognitive development a good predictor of behavior in the tutoring system?

Our objective was to predict student's behavior in the tutor. In this section, we examine the behavior of students with different cognitive levels in predicting success with whole number problems and fraction problems.

4.1 Behavior in whole number problems

When a session in MFD starts, the student first goes through a section of problems about whole numbers (addition, subtraction, multiplication and division). The average amount of time spent per student per whole number problem was considered in a correlation analysis against students with different cognitive levels.

There is a significant correlation (Pearson two-tailed, R=-0.391, p=0.006) that shows that children with lower cognitive levels spend more time solving whole number problems. This suggests that students with higher cognitive levels are faster solvers of whole number problems, for both the experimental group (students who received help) and control group (students who did not receive help). Figure 2 shows the relationship between time spent in whole number problems and cognitive level.

Because total time spent on the tasks might not be a very strong predictor of performance (because some students might be intrinsically slower –perhaps more reflective– workers than others), we decided to investigate an alternative measure of "speed". We looked at how many problems students at different cognitive levels needed to reach mastery of whole numbers. Mastery of whole numbers is considered to be reached when the student solves a certain number of problems for each whole number operation (+, -, x, /) with little or no help at all. The result was a significant correlation between these two variables (Pearson two-tailed, R=-0.39, p=0.007). Figure 2 also shows the relationship between cognitive level and number of whole number problems seen.

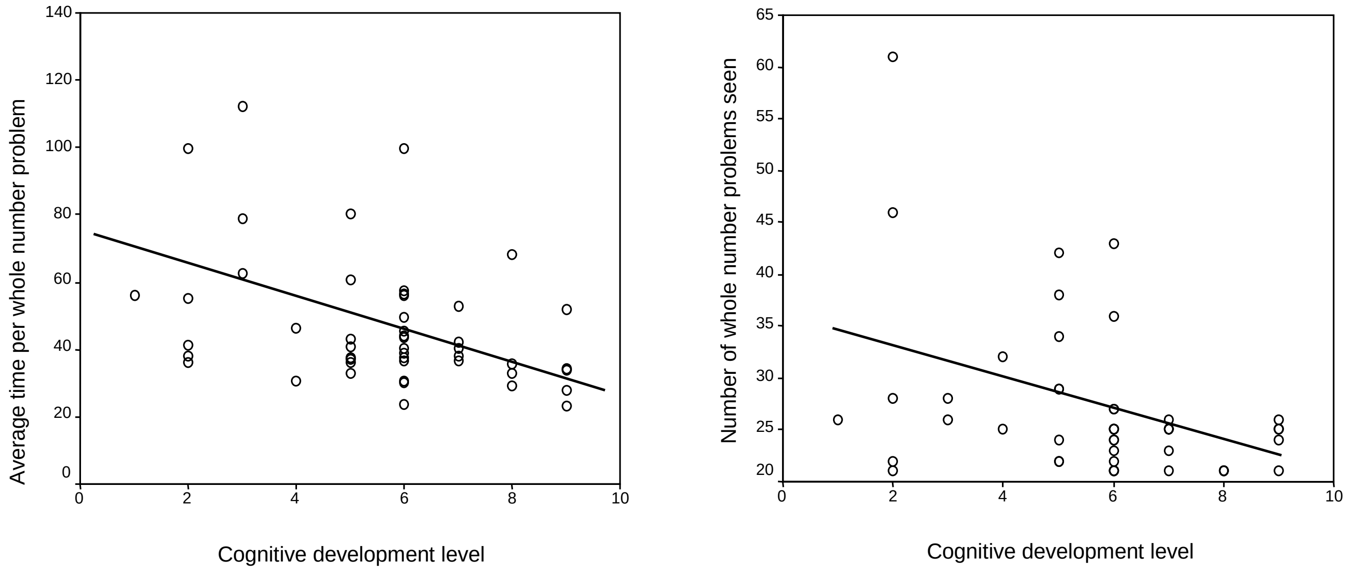

Figure 2: Average time spent in whole number problems for students with different cognitive levels (left). Total number of problems needed to reach mastery of whole numbers for students with different cognitive levels (right).

Students with low cognitive levels needed more problems on average to reach mastery of whole number skills than students with high levels. To verify that this was true (because there was a high variance for students in the lower levels), we performed an independent samples t-test to compare the number of problems required by students above and below a median cognitive level. The means of these two groups were significantly different (two tailed t-test, p=0.004). We also changed the low level and high level groups by pushing the limit between them back and forth, to make sure that it was not a special limit value that created two significantly different high and low level groups. The significance between the two groups remained despite these changes. Table 3 shows the differences between the two groups. The limit between them was, in this case, at a cognitive level of 5.

	N	Mean # problems	Std. Dev.	Std. Error Mean
High level students	28	24.7500	4.7190	.8918
High level students	28	24.7500	4.7190	.8918
Low level students	18	30.8333	10.3142	2.4311
Low level students	18	30.8333	10.3142	2.4311

Table 3: Total number of problems required to reach mastery of whole numbers for students with different cognitive levels

In general, the only students who needed to see many problems to master whole number skills were those with very low cognitive levels. Meanwhile, if the student had a high cognitive level, it was guaranteed that few problems would be enough to master whole number skills.

4.2 Behavior in fraction problems

The tutor determines the type and difficulty of problems given to students. It will move students on to the fraction section only when they have shown mastery of whole number problems. We are curious to know if the hints had been appropriate for students at any cognitive level or at some level in particular. We verified this issue by comparing the performance of those students who did not receive any intelligent help against the performance of those students who were provided the tutor’s help. This will tell us how good the fractions’ hints were for students with different levels of cognitive development. We want to test this particularly for the fraction section because the hints given for fraction problems were much stronger than those given for whole numbers, which provided non significant differences in behavior between the control and experimental groups. We will not measure performance this time as the number of problems that the student needed to reach a certain mastery level. This cannot be done because many students finished the last session in the tutor without reaching mastery for the entire fraction section.

Thus, performance will be measured as the number of actual problems solved weighted by the difficulty of those problems. For example, problems that use operands with unlike denominators are more difficult to solve than those with like denominators. The number of sub-skills that are involved in solving the problem determine the difficulty level of a problem. Finding a common denominator, adding numerators, finding equivalent fractions and simplifying are examples of sub-skills.

We found a significant positive correlation between cognitive development level and performance for those students who had not received the intelligent help (Pearson two-tailed, R=.584, p=0.007). These results show that cognitive level is directly related with performance in the fraction problems. This relationship is not seen for the experimental group, who received intelligent help. This effect can be explained by the fact that when there is no intelligence in the tutor, performance depends on the capabilities of the student. It also means that the current hints seem to be best designed for a group of students with middle level of cognitive development. Furthermore, it means that intelligence in the tutor helped students of average cognitive ability –Piagetian levels 4 to 6, which is late concrete operational stage—to move to a higher performance level. Figure 3 shows these results.

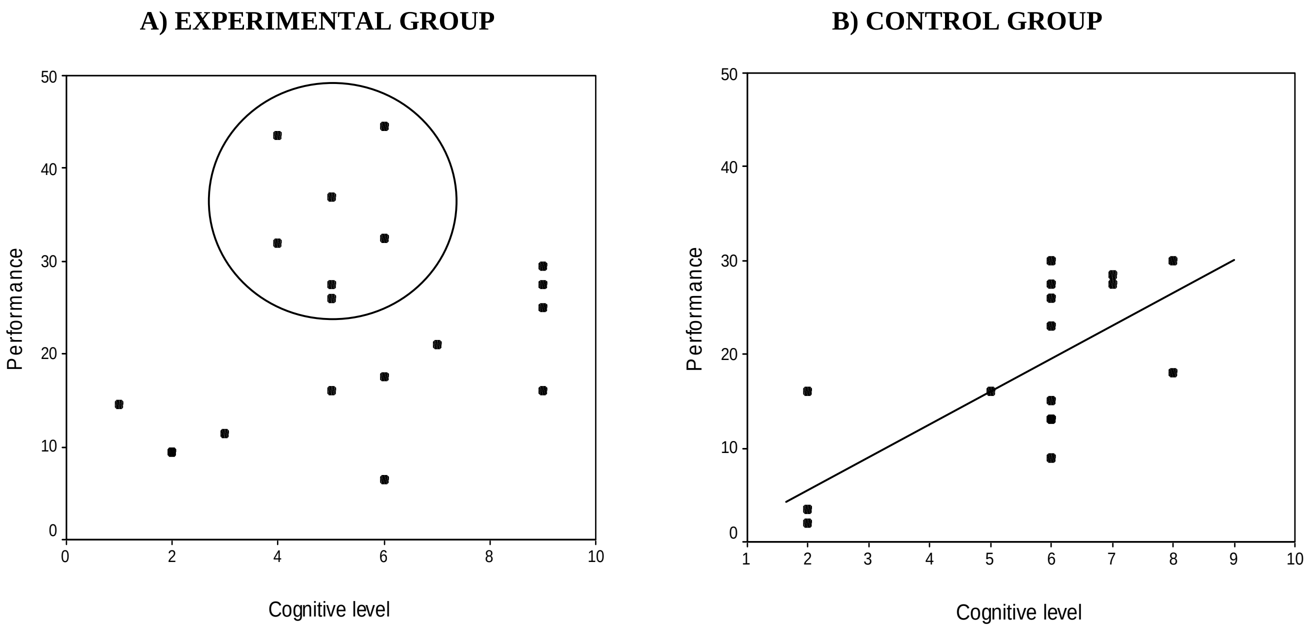

Figure 3: Relationship between cognitive level and performance for the fraction problems.

5. Why didn't the hints seem to help the low cognitive level and high cognitive level students?

The hints in the fractions section showed certain features that prevented a low-level student (cognitive level < 4) from improving her performance. Our hypothesis is that those students didn't understand some of the given hints. The results for the Piagetian experiments for the low-level group are the following:

- 29% failed at number conservation.
- 0% failed at serialization
- 86% failed at reciprocity and reversibility
- 100% failed at area conservation, class inclusion, functionality, proportionality, experiment design and combinatorial analysis.

Two reasons make us believe that the hints didn't work for the lower level students. One of them is that the hints involved higher levels of abstraction than the level these students could handle. Explanations were either highly numeric, or they showed abstract visual representations of fractions in the form of bars. We think that these students need explanations that would connect these concepts to real life (pizzas, dinosaurs, fish, etc.) before encouraging them to transfer this knowledge to representations that are more abstract. Another reason is that they were supposed to have spatial abilities to understand the space transformations involved in the hints. This can be seen in figure 4 where students were supposed to be aided in finding equivalences between fractions. In many cases, a bar divided in fourths was now divided in eighths, but the painted area remained the same. Students were supposed to understand that this transformation made a new fraction that was equivalent to the previous one. However, this operation involves mental manipulation of space to understand that the areas were equivalent although the divisions had changed. The failure at this ability can be understood if we take into account that 100% of these students failed at the area conservation task, were they had to use a similar concept. In addition, failure at reversibility and reciprocity shows these students' poor ability to reason with space.

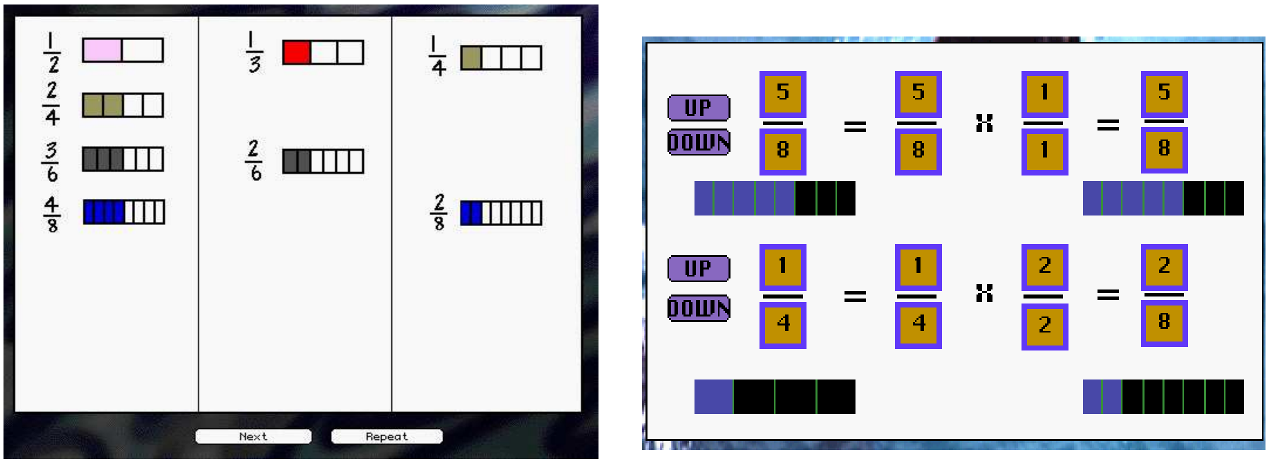

Figure 4: A hint and a widget to scaffold students on equivalence of fractions

An interesting question is why the hints for students at higher levels (cognitive level > 6) seemed to be useless. These students should have had the capacity to understand these hints. We think that hints in fact made a difference for this group of students. We compared the rate of failure for students with high cognitive levels in the experimental and control groups. A student is considered to fail in a problem if he has entered a wrong answer a certain number of times. The response of the tutor is to give up on that problem and propose a new one. We found out that those high-level students who did not receive hints had a significantly higher failure rate (two-tailed independent samples t-test, $p=0.021$). This means that the hints made a difference for those high-level students who received hints. Thus, if the hints were useful, then the students were able to understand the hints. This is important – even if both groups achieved the same performance- because the experimental group was spending time in learning how to solve problems. Meanwhile, the control group was spending time in trying possible answers until some one succeeded. The experimental group was spending time in the hints, which actually delayed them, but also empowered them to answer questions correctly. We think that if the students kept on using the tutor, the experimental group would not have needed the hints any more and would have outperformed the control group. This difference in failure rate also stands for the medium cognitive level group (two-tailed independent samples t-test, $p=0.028$).

Another interesting feature about this group (high level of cognitive development, experimental group) is that they reached a reasonably high level of performance by going through a very low number of problems. The number of problems seen by this group was significantly lower than those seen by a subset of the middle level group which achieved the same performance (two-tailed independent samples t-test, $p=0.05$, performance level between 15 and 35). This suggests that people in the middle level failed more problems, so that they needed to go through more problems to get the same performance. Maybe this group of middle-level students was more focused in the system than the high level students, and that is why they had more time to go through more problems. This hypothesis was confirmed when we analyzed the failure rate. The high-level group had a significantly lower failure rate than this medium level group (two-tailed independent samples t-test, $p=0.017$).

Even more interestingly, the number of problems that the experimental high-level group saw was not significantly different with respect to the number of problems seen by the experimental low-level group. Actually, it was similarly low. This means that both groups were spending the same average time in one problem. However, because the high-level students were succeeding in those problems and the low-level group was failing to give correct answers, their performance level was higher.

To sum up, we know that students with higher cognitive levels can handle higher levels of abstraction (Ginsburg et al., 1988), while lower level students cannot. The hints that we presented were too abstract for the low-level group, so that they were not useful at all for this group of students. Therefore, we believe students with low levels of cognitive development could benefit from concrete (visual, manipulative, with real-life objects) hints, while students with higher levels could benefit from hints that are more abstract. However, this is a hypothesis for the time being. We need to test this hypothesis, and to be specific about what "concrete" and "abstract" hints mean in practice. Our next step will be to generate different kinds of hints with different features, and to test which ones are more effective for students at different cognitive levels. The degree of manipulation (clicking, dragging, etc.), the amount of information, the amount of numerical symbolism and the degree of freedom given to the student are examples of these features. In addition, hints could have generic cognitive pre-requisites (reversibility, proportionality, etc.) that the student should demonstrate before a certain hint is presented. Then, hints could be selected according to the student's cognitive skills as measured by our Piagetian test.

We would like to test the hypothesis that students with low cognitive levels need different kinds of hints by building an experiment where students are semi-randomly given hints with different levels of information. We plan to measure how appropriate each hint is through both statistical analysis and machine learning. However, we still need to establish how to measure the "appropriateness" of a hint. We are considering two possible approaches. The first is to take into account the average time from the moment the person sees the hint until the moment she enters the correct answer. The second is to consider the number of mistakes made after receiving the hint and before the correct answer is entered.

6. Conclusions and future work

We have constructed a test to measure elementary school students' level of cognitive development according to Piaget's theory of developmental stages. We have adapted classic tasks used to measure these levels for use on computer. The test requires approximately 10 to 15 minutes for students to complete. This measure predicts student performance at a variety of grain sizes: effectiveness of hints received, rate of failure, amount of time to solve problems and the number of problems students need to attempt to master a topic.

The data we have obtained from 46 sixth grade students strongly suggests that cognitive level is a useful variable to add to an intelligent tutoring system, when the population of students is around 10 years old. These results are similar to prior predictive work in the field (Anderson, 1993; Shute, 1995). However, our measure takes little time to administer, which is an advantage given the relatively brief period of time most tutors are used.

We plan to pursue this research along several independent paths. First, we are interested in improving the instrument itself. Based on expert assessment, and the high correlation of our two test scores, it is likely the pretest has measured the construct in which we are interested. However, from observing students it is clear that some of the Piagetian tasks are either confusing to some students or that some students are answering them differently than we expected. It was not easy to replace the human interviewer by the computer and still get the same answers. We are therefore refining the pretest questions. This revised instrument will be tested in February 1999 and May 1999.

Another path is augmenting the tutoring knowledge by including Piagetian information about each hint. The tutor can use this knowledge to avoid presenting hints that are beyond the student's understanding. Finally, we are determining how to add cognitive development to the tutor's teaching and update rules. This is difficult, as most teachers/tutors do not think about this information when instructing. Therefore, we are considering using machine learning techniques (Stern et. al, 1999) to allow the tutor to determine for itself how to best use this information. We are also planning to give the students a paper and pencil pre-test and post-test with similar problems than those given by the tutor to test how much teaching the hints provided.

Acknowledgements: We acknowledge support for this work from the National Science Foundation, HRD-9714757. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting agency.

Anderson, J. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Beck, J.; Stern, M.; Woolf, B. (1997). Using the student model to control problem difficulty. *The Sixth International Conference on User Modeling.*

Bukatko, D.; Daheler, M. (1995) *Child Development: A Thematic Approach*. Houghton Mifflin Co.

Ginsburg, H,; Opper, S. (1988). *Piaget's Theory of Intellectual Development*.

Mayer, Richard E. (1977). *Thinking and problem solving: an introduction to human cognition and learning*.

Piaget, J. (November 1953) How Children Form Mathematical Concepts. *In Scientific American*.

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

Shute, V. (1995). Smart evaluation: Cognitive diagnosis, mastery learning and remediation. In *Proceedings of Artificial Intelligence in Education.* Pages 123—130.

Stern, M.; Beck, J.; Woolf B. (1999). Naïve Bayes Classifiers for User Modeling. Submitted to *The Seventh International Conference on User Modeling.*

Voyat Gilbert E. (1982). *Piaget Systematized*.