A useful prediction variable for student models: cognitive development level

Ivon Arroyo, Joseph E. Beck, Klaus Schultz, Beverly Park Woolf

Computer Science Department and School of Education, University of Massachusetts, Amherst

Making a realistic update of a user model based on evidence in the environment is not an easy task, unless a great deal of time with a large variety of users is available. Creating general categories of users that behave in a certain way is important for any kind of user model. To obtain such a broad classification we need to understand general factors that influence user behavior. We describe the use of pretests to measure cognitive development of student users and how this factor is input to a student model. We describe how measures of cognitive ability enhance the predictive power of a student model in an intelligent tutoring system for a population of young (elementary school) students.

Keywords: Cognitive resources, Piaget, student modeling.

Ivon Arroyo
Ivon@cs.umass.edu
Computer Science Department, University of Massachusetts, Amherst, MA 01003
Telephone: (413) 545-0582
Fax: (413) 543-1249

1.Introduction

In this paper we describe the use of Piaget's notion of cognitive development to improve a tutor's reasoning ability (Piaget, 1953). We are interested in finding information that not only predicts a student's overall performance, but that can also be easily applied to actual tutoring decisions. There has been some prior work related to this in user modeling. A factor analysis with data from the LISP tutor (Anderson, 1993) demonstrated that there were two factors that were useful at explaining student performance. These factors could be described as acquisition of new information, and the student's ability to retain old knowledge. Stat Lady (Shute, 1995) found that a six-hour pretest was predictive of student learning. Unfortunately, none of the measures used for adults is useful for reasoning about young students. There have been some attempts at finding "low cost" metrics to predict student performance. Other work on the LISP tutor (Anderson, 1993) demonstrated a correlation between students' performance and their math SAT scores. Work on a mathematics tutor (Beck et al., 1997) attempted to derive online measures of acquisition and retention for each student as she was interacting with the tutor. Regretfully, this work was limited in that it was difficult to see how these values were properly measuring acquisition and retention, and how good it was to incorporate them as learning rates. Our current work, by building on an established theory of cognition, should be easier to apply.

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 a subset of these topics: 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 intended to determine if the students were at one of the last two stages of cognitive development proposed by Piaget –concrete operational stage and formal operational stage–. [2]

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). These tasks 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.

- *Substance conservation:* 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 asked students to determine where the level of water was going to be in the empty vessels if the liquid in two identical vessels was poured into them.

- *Area conservation:* Students were asked to compare two areas of the same size but different shape.

---

[1] Due to absentees among students, we only have complete data for 46 students.
[2] 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–. In fact, the categorization of cognitive development into two discrete stages is an oversimplification of the complicated story of intellectual development. 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.

- *Serialization*: Students had to order a group of pencils from the shortest to the longest one.

- *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.

- *Functionality*: Students had to invent an algorithm to solve a problem of ordering 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.

- *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 new measurement system.

- *Combinatorial analysis*: Students were asked to generate combinations of four switches to open a safe.

3. Description of results

The number of Piagetian tasks that the student accomplished was used as a measure of cognitive development. The mean number of correct answers for the sixth grade pupils in the study was 5.7 (out of 10), with a standard deviation of 2.1. Most students could do approximately half of the tasks correctly. 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).

We also considered another measure of cognitive development, which weighted the tasks according to their expected difficulty. Thus, students who had succeeded on a relatively difficult experiment (like combinatorial analysis) would be considered to have a higher cognitive level than those who had succeeded on a less difficult task (like number conservation). The two measures of cognitive development turned out to be highly correlated. This confirms our

hypothesis about the relative difficulty of the tasks: when students succeeded at only a few experiments, they tended to be the ones that we considered easiest (Pearson two-tailed, R=0.975, p=0.000).

4. Is cognitive development a good predictor of performance?

Our objective was to predict student's performance at a variety of tasks. 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 Relationship to time spent 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.
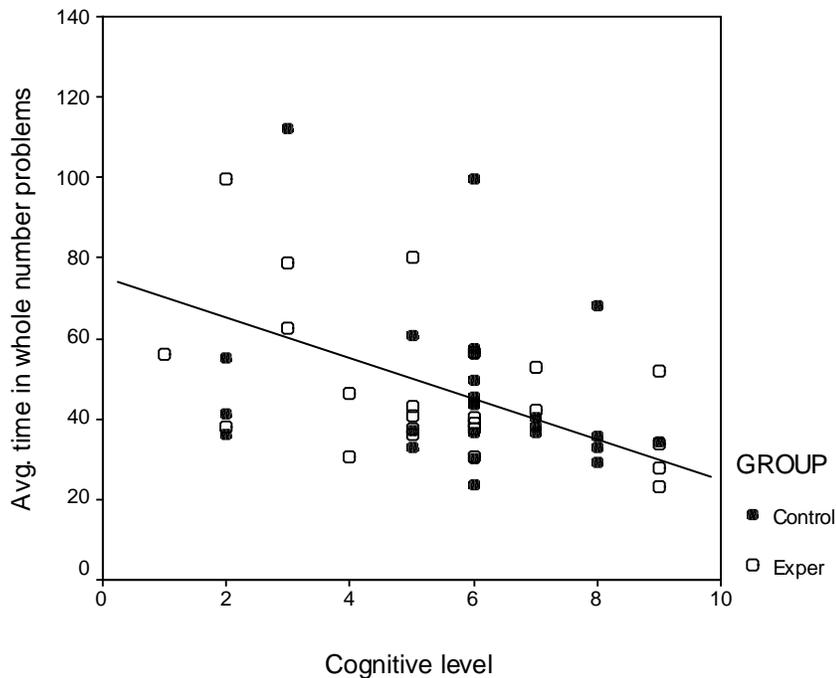

Figure 1: Average time spent in whole number problems for students with different cognitive levels

There is a significant correlation that shows that children with lower cognitive levels spend more time solving whole number problems (Pearson two-tailed, R=-0.384, p=0.007). 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 1 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=-390, p=0.007). Figure 2 shows the relationship between cognitive level and number of whole number problems seen.
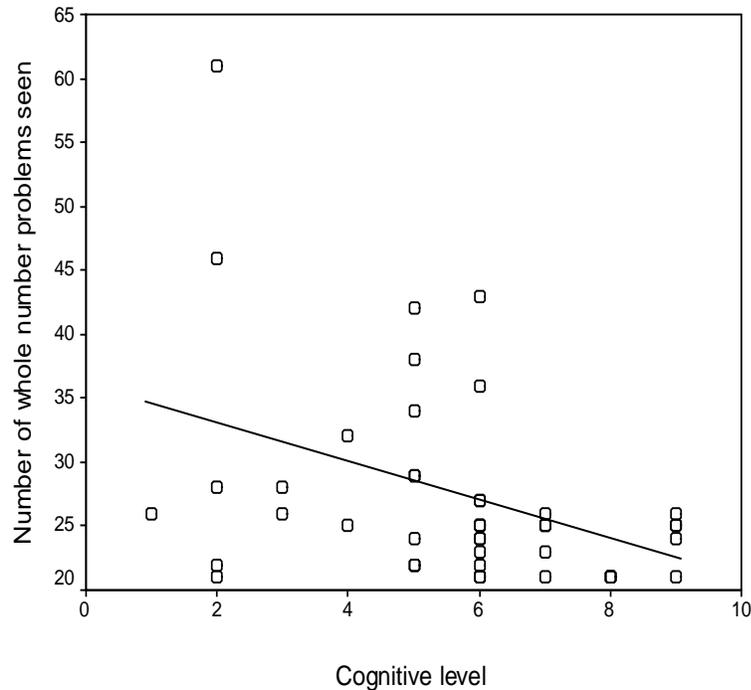

Figure 2: Total number of problems needed to reach mastery of whole numbers for students with different cognitive levels

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 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 just a special limit value that created two different high and low level groups. The significance between the two groups remained despite these changes. Table 1 and figure 3 show the differences between the two groups. The limit between the two groups was at a cognitive level of 5.

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

Table 1: Total number of problems required to reach mastery of whole numbers for students with different cognitive levels
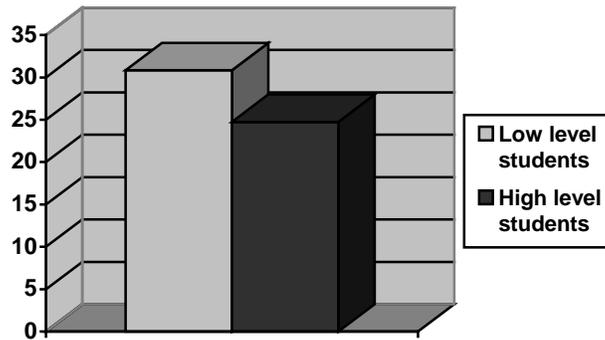
Figure 3: Average number of whole number problems that the students required to reach mastery of whole number problems

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 Relationship to performance in fraction problems

The tutor determines the type and difficulty of problems generated for students. It will move students on to the fraction section only when they have shown mastery of whole number problems. We are particularly interested in determining how the level of cognitive development is related to student performance in the fraction section of the tutor for those students who did not receive any intelligent help from the tutor and compare it against the performance of those students who were provided the tutor's help. This will tell us how good the hints were for students with different levels of cognitive development. We want to test this 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 cannot measure performance as the number of problems that the student required to reach a certain mastery level because all the students finished the last session in the tutor at different levels (without reaching mastery for the whole section). Thus, performance will be measured as the number of actual problems solved weighted by the difficulty of those problems. We decided to use this measurement of performance because there are many difficulty levels of problems. For example, problems that use operands with different denominators are more difficult to solve than those with same 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 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 is being used (see figure 4). This effect could 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 are 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.

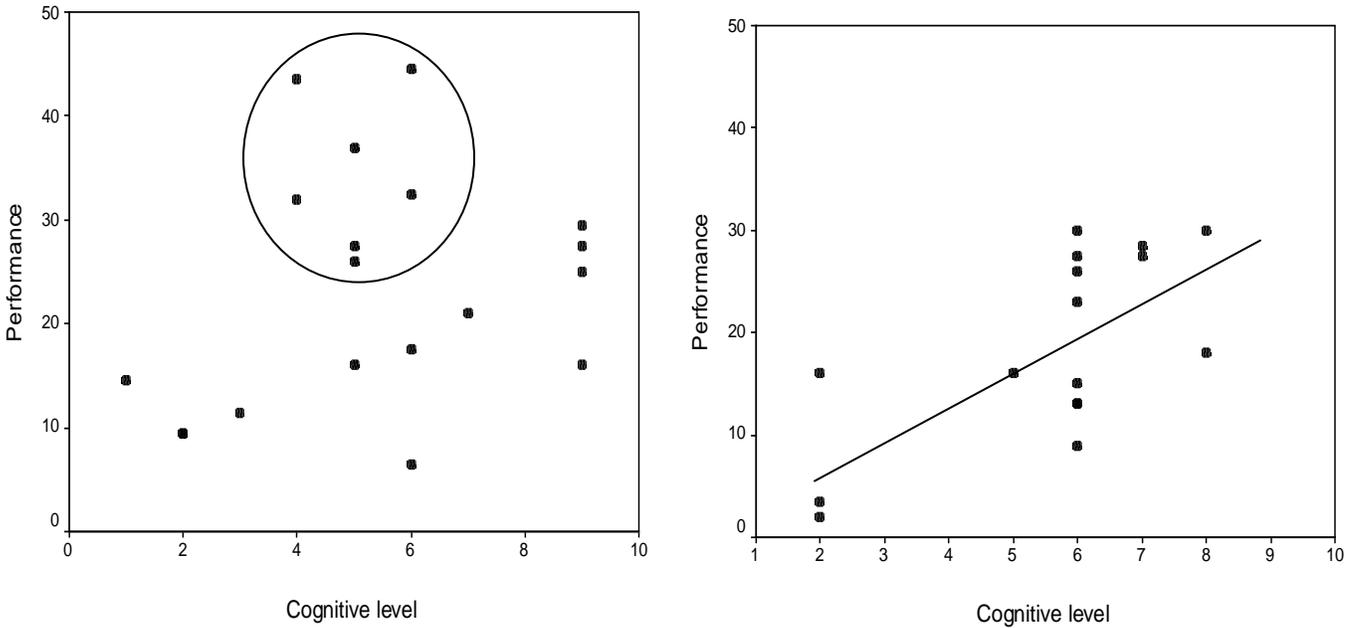**A) EXPERIMENTAL GROUP**   **B) CONTROL GROUP**

Figure 4: Relationship between cognitive level and performance for the fraction problems

Two clusters can be identified if we overlap graphs 4A and 4B, regardless of the student being in the control or experimental group (i.e. regardless of how much intelligent help they got). A low cognitive level group (less than 4 correctly solved tasks) can be detected that shows low performance (mean=9.19, std. deviation = 4.91) and a medium-high cognitive level group (more than or equal to 4) that achieves a higher performance (mean=25.37, std. deviation = 11.07).
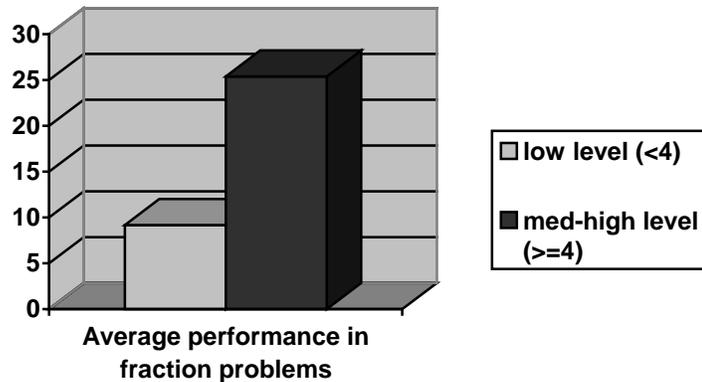

Figure 4: Differences in performance between a low and a high cognitive level groups

This difference is significant (independent samples two-tailed t-test, p=0.001) and suggests that students at low levels of cognitive development (early concrete operational stage) are slower learners, and that the current hints were apparently not well designed for that group of students.

5. How can determination of cognitive level be useful in intelligent tutoring systems?

We have shown that knowledge of a students' cognitive level is a valid predictor of his performance in using an intelligent tutoring system. Now we will consider how an ITS can use this variable to enhance its teaching.

5.1 Student promotion

As it has been shown in section 3.1, students with low levels of cognitive development require a larger number of problems to master skills than high level students. The tutor can use cognitive level as speed of learning parameter that influences the number of problems the student must solve before the tutor believes he has mastered the skill in question. For this purpose, we can find a best-fit curve that determines the number of problems needed to reach mastery as a function of cognitive development. We can also extract a learning rate by obtaining probabilities of the student knowing a skill given that she has gone through a certain number of problems. This can be useful for student models using Bayesian networks.

5.2 Confusion Detection

Students with high cognitive levels take less time to solve problems. Thus, if a student with high level of cognitive development is taking a long time to solve a problem, then there may be something wrong with the student's understanding of the problem. This could be a good opportunity to initiate some hint suggesting she re-read the problem to make sure that there is not a misunderstanding in what the student is required to do. This is in contrast to a student with a very low cognitive development taking a long time on such a problem. It is more likely that such a student is having difficulty solving the problem, so the feedback should differ.

5.3 Hint selection

Students with higher cognitive levels can handle higher levels of abstraction (Ginsburg et. al, 1988), while lower level students cannot. Therefore, we believe students with low levels could benefit from more concrete (visual, manipulative) hints, while student with high levels could benefit from more abstract (symbolic, with use of generalizations) hints. 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 text, 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.

5.4 Number of hints and granularity level

Because low cognitive level students need more help, we will use cognitive level to vary both the number of hints given to low cognitive level students, and the amount of information provided in each hint. We would like to test the hypothesis that students with low cognitive levels need more information in each hint by building an experiment where students are semi-randomly given hints with different levels of information. We would then be able to determine how much hinting students with different cognitive levels need, given that they have a certain level of mastery of the skill in question.

We plan to establish 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: number of hints received, 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 a student model in 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. 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.

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. and Beck, J. (1999). Naïve Bayes Classifiers for User Modeling. Submitted to *The Seventh International Conference on User Modeling*.

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