Testimony to the National Mathematics Panel

Steven Ritter  
Carnegie Learning, Inc.

John Anderson  
Carnegie Mellon University

November 6, 2006

Cognitive Tutor: Tracking learning in real time

Thank you very much for inviting us to present information about how, in our development of the Cognitive Tutor® curricula, we have applied basic research to the improvement of mathematics education.

In these remarks, we will provide some background about Carnegie Learning and then explain what we consider scientifically-based research to involve, how we have applied research in the construction of our curricula and what we view as the potential of technology to dramatically improve mathematics education in the United States.

Carnegie Learning Background

The work that led to Carnegie Learning’s Cognitive Tutors began in the psychology and computer science departments at Carnegie Mellon University. John Anderson had been developing the ACT (later ACT-R) theory of cognition. ACT-R is a Unified Theory of Cognition (Newell, 1973, 1990) that aims to explain the full range of human cognition. ACT-R was implemented as a computer program, which has the advantage of requiring the theory to be precise about all of its claims. Anderson had seen great success in using ACT-R to model laboratory results in learning, memory and problem-solving (c.f. Anderson, 1983), and he was challenged to show that the same basic approach could explain cognition outside of a laboratory environment.

In its application to psychological laboratory studies, the aim of an ACT-R model is to predict behavior. In order to predict behavior, the model needs to correctly represent human knowledge and also understand how that knowledge results in particular behaviors. Applied to education, this representation of knowledge results in predictions about what students can and cannot do as well as predictions about what activities and experiences will help students learn to achieve curricular goals.

The representation of knowledge inherent in this kind of model is called a “cognitive model,” and the approach of using a cognitive model in a tutoring system has come to be called a “Cognitive Tutor.” The first tutoring systems built in this way addressed computer programming and mathematics (Anderson, Boyle, Farrell & Reiser, 1987; Anderson, Conrad, & Corbett, 1989; Anderson, Boyle, Corbett & Lewis, 1990).

ANGLE, a geometry proof tutor (Koedinger and Anderson, 1993) was successful in a school setting. Its success, however, appeared to be highly dependent on the teacher’s ability to integrate the tutoring software into broader classroom goals. This, along with Koedinger’s personal experience teaching a geometry class, focused the research group on the importance of working with teachers and administrators to understand schools’ curricular needs more broadly. As a consequence, the research team...
for the products that became Carnegie Learning’s Cognitive Tutors included Bill Hadley, who had taught
mathematics for almost 30 years and was the recipient of the Presidential Award for Mathematics
Teaching. This team set out to build curricula that were based in solid cognitive research, were focused on
emerging national and state standards and that addressed the practical needs of students, teachers and
administrators.

One decision was to include textbooks in addition to the software. The inclusion of text allowed the
curricula to include some aspects (such as collaboration, diagramming and writing about mathematics)
that were easier to do on paper than on the computer. The combination of text and software also helped to
position the software as a regular, routine part of the mathematics instruction. Instead of using the
software as a “bonus” for advanced students or as a review for students who were lagging, the hybrid
curricula set the expectation that software could be used as part of the primary instruction. Pilot
implementations led to a model where students used the software two days per week, with classroom
activities structured by the text the other three days each week.

The curricula proved to be educationally successful (Koedinger, Anderson, Hadley and Mark, 1997;
Ritter and Anderson, 1995; Koedinger, Corbett, Ritter & Shapiro, 2000) and popular with students and
teachers.

**Scientifically-based research and development**

In our view, scientifically-based research involves more than simply the demonstration that a curriculum
is effective. An essential component is understanding why the curriculum is effective. Without a
theoretical framework as a guide to understanding the conditions that lead to effective mathematics
instruction within a curriculum, we have little hope of expanding and improving instruction over time.

We think of the process of building a curriculum as having four parts:

- Having a solid theoretical basis
- Applying the basic theory to the particular domain and objectives of interest
- Evaluating results
- Developing and implementing a methodology for improving the curriculum, based on use

**Theoretical Basis**

ACT-R (Anderson, 1990, 1993; Anderson & Lebiere, 1998; Anderson, Bothell, Byrne, Douglass, Lebiere
& Qin, 2004) forms the primary theoretical basis of the Cognitive Tutors. The primary use of the ACT-R
theory has been to reproduce important characteristics of human behavior, including error patterns and
response times. Most of this work has been conducted in the laboratory, but ACT-R has also been applied
outside of the laboratory in areas related to human-computer interaction, training and education. This
work has resulted in hundreds of publications (see [http://act-r.psy.cmu.edu/publications/index.php](http://act-r.psy.cmu.edu/publications/index.php)).

A full explanation of ACT-R is beyond the scope of this testimony, but some of the tenets important to
education (Anderson, 2002) include:

- There are two basic types of knowledge: procedural and declarative. Declarative knowledge
  includes facts, images and sounds. Procedural knowledge is an understanding of how to do things. All
  tasks involve a combination of the two types of knowledge. As we learn, we generally start out with
  declarative knowledge, which becomes proceduralized through practice. Procedural knowledge tends
to be more fluent and automatic. Elements of procedural knowledge are referred to as “rules” or
“productions” because they specify the conditions under which they are applicable and the actions (including changes in mental state) that result from applying them. Declarative knowledge tends to be more flexible and also more broadly applicable than procedural knowledge. We often refer to elements of declarative knowledge as “facts.”

- The knowledge required to accomplish complex tasks can be described as the set of declarative and procedural knowledge components relevant to the task.

- Both declarative and procedural knowledge become strengthened with use (and weakened with disuse). Strong knowledge can be remembered and called to attention rapidly and with some certainty. Weak knowledge may be slow, effortful or impossible to retrieve. Different knowledge components may represent different strategies or methods for accomplishing a task (including incorrect strategies or methods). The relative strength of these components helps determine which strategy is used.

- Learning involves the development and strengthening of correct, efficient and appropriate knowledge components.

It is important to understand that our terminology here differs somewhat from the same terms as they are used in an educational context. For example, a “procedure” in ACT-R is simply a component of knowledge that can produce other knowledge components and/or lead to external behavior. In mathematics education, we might refer to the “procedure” of solving a linear equation. An ACT-R model of that task would consist of many productions and facts that are brought to bear. Even a simple task like adding integers may consist of many productions, including ones associated with recalling arithmetic facts, executing counting actions, etc. (c.f. Lebiere, 1999).

This view that emerges from ACT-R is that learning is a process of encoding, strengthening and proceduralizing knowledge. This process happens gradually. New knowledge will be forgotten (or remain weak enough to stay unused) if it is not practiced, and elements of knowledge compete to be used, based on their strength (Siegler & Shipley, 1995). Since the ability to perform a task just relies on the individual knowledge components required for that task, education will be most efficient when it focuses students most directly on the individual knowledge components that have relatively low strength.

The interaction between declarative and procedural knowledge leads to an emphasis on active engagement with the conceptual underpinnings of procedures, so that students appropriately generalize this knowledge (Rittle-Johnson & Siegler, 1998; Rittle-Johnson, Siegler & Alibali, 2001; Rittle-Johnson & Koedinger, 2002; 2005). Since procedural knowledge includes the context in which it is applicable, educational activities need to be structured such that students are able to practice procedures within an appropriate range of contexts.

This decomposition of complex tasks into individual knowledge components leads to a pedagogical model emphasizing practice of individual components, independent of the larger task. At the same time, we need to recognize that some knowledge components that are inherent to the larger task (such as integration of information from different smaller components), which provides another rationale for emphasizing performance within an appropriate context.

To use a sports analogy, it is important for batters to take batting practice, because this will allow a baseball player to receive intensive practice with most of the skills involved with hitting a ball. However, it is also important for the batter to play in games, since some skills (such as reading the infield) will only come up in that context.
Application of principles

Although the ACT-R theory indicates the basic pedagogical strategies likely to be effective in instruction, it does not specify the particular skills that comprise the ability to, for example, solve a linear equation. In order to create instruction in mathematics, we need to understand the knowledge components involved in completing a particular task. It is not enough to know the components involved in expert performance of the task; we also need to know the components exercised by students learning to perform the task. Much of our applied research in mathematics has concerned identifying the particular skills and methods that students use to complete mathematical tasks (c.f. Corbett, McLaughlin, Scarpinnatto and Hadley, 2000; Mark and Koedinger, 1999; Koedinger and Anderson, 1990). Often these skills do not correspond to expert beliefs (Koedinger & Nathan, 2004; Nathan & Koedinger, 2000a; 2000b).

One technique that we have used to understand how student approach mathematics problems is to track their eye movements as they work through a problem (Gluck, 1999). Consider the task of a student completing a table of values based on a word problem like that shown in Figure 1.

In Figure 1, the student has completed part of the table corresponding to the word problem, including the column headings, units of measure expression and the number of hours asked for in the first question. The student next needs to calculate the amount of money remaining after two hours. The student might perform the task in at least two ways. First, they might reason from the problem scenario (perhaps imagining having $20 and then using repeated subtraction to calculate the money left after spending $4 two times). The second method would be to use the algebraic expression and then substitute 2 for X and calculate the result. If a student has produced the table shown in Figure 1 (including writing the algebraic expression for the amount of money left), we might expect that students would then use the algebraic expression and execute the second method. In fact, Gluck found that, about 13% of the time, when students were answering a question like question 1, they looked at the problem scenario but not the expression. 54% of the time, students looked at the expression (sometimes along with the scenario). Almost 34% of the time, they looked in neither place.

As a result of this and other data (c.f. Koedinger & Anderson, 1998), the Cognitive Tutor curriculum treats finding the algebraic expression for simple word problems as an induction task. Students are asked to solve the individual questions (“How much money would be left after 2 hours?” and “When will you run out of money?”) first, and then use a generalization of their reasoning to come up with the algebraic expression. In later units of curriculum, as the situations and algebraic expressions become more...
complex, we focus students on going from the word problem to the expression and then using the expression to compute specific values.

Beyond the design of mathematical tasks, the ACT-R theory guides instruction in the Cognitive Tutor because the software includes an active cognitive model, which is similar to an ACT-R model within the software (Corbett, Koedinger & Anderson, 1997). This model serves two purposes. First, the model follows student actions in order to determine the particular student’s strategy in solving a problem. The technique by which it does this is called model tracing. Second, each action that the student takes is associated with one of more skills, which are references to knowledge components in the cognitive model. Individual students’ performance on these skills is tracked over time (and displayed to students in the “skillometer”). The Cognitive Tutor uses each student’s skill profile to pick problems that emphasize the skills on which the student is weakest (Corbett and Anderson, 1995). In addition the skill model is used to implement mastery learning. When all skills in a section of the curriculum are determined to be sufficiently mastered, the student moves on to the next section of curriculum, which introduces new skills.

**Careful Evaluations**

The development of curriculum involves many decisions, and there is often room for disagreement about how learning theory should be applied in particular cases. For that reason, we believe that careful evaluation is an essential part of the process.

Our development process has included many formative evaluations of individual units of instruction (e.g. Aleven & Koedinger, 2002; Corbett, Trask, Scarpinatto & Hadley, 1998; Koedinger & Anderson, 1998; Ritter and Anderson, 1995). In addition, we have conducted several large evaluations of the entire curriculum (taking text, software and training components as a single manipulation).

Early evaluations of Cognitive Tutors for programming and geometry showed great promise, with effect sizes of approximately 1 standard deviation (Anderson, Corbett, Koedinger & Pelletier, 1995). In studies of the Algebra I Cognitive Tutor conducted in Pittsburgh and Milwaukee (Koedinger, Anderson, Hadley and Mark, 1997), students were tested both on standardized tests (SAT and Iowa) as well as performance-based problem-solving. Cognitive Tutor students significantly outscored their peers on the standardized tests (by about 0.3 standard deviations), but the difference in performance was particularly pronounced on tests of problem-solving and multiple representations, where the Cognitive Tutor students outscored their peers by 85%, representing effect sizes from 0.7 to 1.2 standard deviations.

In Moore, OK, a study was conducted where teachers were asked to teach some of their classes using Cognitive Tutor and some using the textbook they had been previously using (Morgan & Ritter, 2002; National Research Council, 2003). The study found that Cognitive Tutor students scored higher on a standardized test (the ETS Algebra I End of Course exam), received higher grades, reported more confidence in their mathematical abilities and were more likely to believe that mathematics will be useful to them outside of school. This study was recognized by the US Department of Education’s What Works Clearinghouse as having met the highest standards of evidence. This study showed effect sizes of approximately 0.4 standard deviations.

The Miami-Dade County school district studied the use of Cognitive Tutor Algebra I in ten high schools. An analysis of over 6000 students taking the 2003 FCAT (state exam) showed that students who used Cognitive Tutor significantly outscored their peers on the exam (Sarkis, 2004). The findings were particularly dramatic for special populations. The study found that 35.7% of Exceptional Student Education students who used Cognitive Tutor passed the FCAT, as compared to only 10.9% of such students using a different curriculum. For students with Limited English Proficiency, 27% of Cognitive
Tutor students passed the FCAT, as opposed to only 18.9% of such students in another curriculum.

**Methodology for Improvement**

ACT-R provides guidelines for educational pedagogy and for constructing tasks that are likely to increase learning. The theory also provides a way for us to test and improve our curriculum over time.

The Cognitive Tutor observes students. As an observer, it sees everything the student does, at approximately 10-second intervals, for two days per week over a school year. However, the cognitive model is not a passive observer. It is continually evaluating the student and predicting what the student knows and does not know. By aggregating these predictions across students, we can test whether the cognitive model is correctly modeling student behavior.

Consider what an observer should see across time in a classroom. If students are learning, they should be making fewer errors over time. However, the activities given to the students over time should also increase in difficulty. In a well-constructed curriculum, these two forces should cancel each other out, leading to a fairly constant error rate over time. In fact, that is what we see in the Cognitive Tutor curricula. Figure 2 shows percent correct, over time, for 88 students using the Cognitive Tutor Geometry curriculum in one school. The percentage correct remains fairly constant over time.

![Percent Correct (all students, all actions)](image)

**Figure 2:** Percent correct over time, considering all student actions in the Geometry curriculum

ACT-R makes the strong claim that learning takes place at the level of the knowledge components. Thus, if we measure percent correct over time, considering only actions that involve a particular knowledge component, we should see an increase (Anderson, Conrad & Corbett, 1989). Figure 3 shows percent correct for the same group of students, considering only those student actions that the cognitive model considers to be relevant to a single skill (calculating the area of a regular polygon, in an orientation where a side is horizontal).
If ACT-R is correct that performance of a complex task is determined by the individual knowledge components contributing to the performance of that task, then each skill in the cognitive model should show a learning curve like this. Failure to see learning on one of the component skills must mean that the cognitive model implemented in the tutor is not correctly representing student knowledge.

In the development of our algebra tutor, we discovered that the model was over-predicting student performance in solving equations of the form $ax=b$. An analysis of the data revealed that the over-prediction was, in part, due to the case where $a=-1$. The explanation for this over-prediction, in retrospect, is obvious. In the case where $a=-1$, the student needs to understand that the expression $-x$ means $-1$ times $x$ and that, otherwise, the equation can be solved using the same operations as would be applied to any equation of the form $ax=b$. (Another way to think about this error is that some students have learned a rule equivalent to “if the equation is of the form $ax=b$, then divide by the number in front of the variable.” But, when the coefficient is $-1$, the student doesn’t see a number, just a negative sign, so the rule does not apply.) Once recognition of $-x$ as $-1$ times $x$ was added to the cognitive model, the Cognitive Tutor automatically adjusts instruction to test whether students have mastered that skill and will automatically provide extra practice on such problems to students who need it. In addition, we can define instruction specifically targeted at this skill.

The process of analyzing learning curves and improving our fit to the data has, to this point, been laborious. We have recently been exploring the possibility of automating the process of discovering flaws in the cognitive model (Cen, Koedinger & Junker, 2005; Junker, Koedinger and Trottini, 2000) and this is an active focus of research at the Pittsburgh Science of Learning Center (www.learnlab.org).

We believe that, in the near future, we will be able to greatly extend our ability to understand and accurately model students’ mathematical cognition. In addition to improved statistical modeling techniques, the expansion of Carnegie Learning’s customer base and the ability to aggregate student data over the internet provides us with the ability to look both more deeply and more broadly at student cognition.

We have now collected data on over 3000 students using the Cognitive Tutor in a pre-algebra class. These data comprise over 8.5 million observations, which amounts to observing an action for each student about every 9.5 seconds. With a database of this size, we expect to be able to detect more subtle factors affecting learning, including the effectiveness of individual tasks, hints and feedback patterns. We are starting to apply microgenetic methods (Siegler and Crowley, 1991) to see if we can identify key learning experiences, which could contribute to cognitive models that better model individual differences in prior
knowledge or learning styles and preferences.

We are optimistic about the potential for Cognitive Tutors to continue to improve in their ability to help students learn mathematics.

References


