

Implications of the ACT-R Learning Theory: No Magic Bullets¹

John R. Anderson
Christian D. Schunn
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

ja@cmu.edu
schunn@gmu.edu

Anderson, J. R., & Schunn, C.D. (2000). In R. Glaser (Ed.), *Advances in instructional psychology (Vol. 5)*. Mahwah, NJ: Erlbaum.

¹ Preparation of this paper was supported by grant N00014-96-1-0135 from the Office of Naval Research.

From Ebbinghaus onward psychology has seen an enormous amount of research invested in the study of learning and memory. This research has produced a steady stream of results and, with a few "mini-revolutions" along the way, a steady increase in our understanding of how knowledge is acquired, retained, retrieved, and utilized. Throughout this history there has been a concern with the relationship of this research to its obvious application to education. The first author has written two textbooks (Anderson, 1995a, 1995b) summarizing some of this research. In both textbooks he has made efforts to identify the implications of this research for education. However, he left both textbooks feeling very dissatisfied -- that the intricacy of research and theory on the psychological side was not showing through in the intricacy of educational application. One finds in psychology many claims of relevance of cognitive psychology research for education. However, these claims are loose and vague and contrast sharply with the crisp theory and results that exist in the field.

To be able to rigorously understand what the implications are of cognitive psychology research one needs a rigorous theory that bridges the gap between the detail of the laboratory experiment and the scale of the educational enterprise. This chapter is based on the ACT-R theory (Anderson, 1993, 1996) which has been able to explain learning in basic psychology experiments and in a number of educational domains. ACT-R has been advertised as a "simple theory of learning and cognition". It proposes that complex cognition is composed of relatively simple knowledge units which are acquired according to relatively simple principles. Human cognition is complex but this complexity reflects complex composition of the basic elements and principles just as a computer can produce complex aggregate behavior from simple computing elements. The ACT-R perspective places a premium on the practice which is required to learn permanently the components of the desired competence. The ACT-R theory claims that to learn a complex competence each component of that competence must be mastered. It is a sharp contrast to many educational claims, supposedly based in cognitive research, that there are moments of insight or transformations when whole knowledge structures become reorganized or learned. In contrast, ACT-R implies that there is no "free lunch" and each piece of knowledge requires its own due of learning. Given the prevalence of the "free lunch myth" we will endeavor to show that it is not true empirically and to explain why it can not be true within the ACT-R theory.

This chapter will have the following organization. First we will describe the ACT-R theory and its learning principles. In the light of this theory, we will identify what we think are the important implications of psychological research for education. We will also address the issue of why so much of the research on learning and memory falls short of significant educational application. We will devote special attention to the issues of insight, learning with understanding, and transfer which are part of the free lunch myth. Finally, we will describe how we have tried to bring the lessons of this analysis to bear in the design of our cognitive tutors (Anderson, Boyle, Corbett, & Lewis, 1990; Anderson, Corbett, Koedinger, & Pelletier, 1995).

The ACT-R Theory

The ACT-R theory admits of three basic binary distinctions. First, there is a distinction between two types of knowledge -- declarative knowledge of facts and procedural knowledge of how to

do various cognitive tasks. Second, there is the distinction between the performance assumptions about how ACT-R deploys what it knows to solve a task and the learning assumptions about how it acquires new knowledge. Third, there is a distinction between the symbolic level in ACT-R which involves discrete knowledge structures and a sub-symbolic level which involves neural-like activation-based processes that determine the availability of these symbolic structures. We will first describe ACT-R at the symbolic level. A symbolic-level analysis of the knowledge structures in a domain corresponds basically to a task analysis of what needs to be learned in that domain. However, as we will see, the availability of these symbolic structures depends critically on the subsymbolic processes.

Declarative and Procedural Knowledge

Declarative knowledge reflects the factual information that a person knows and can report. According to ACT-R declarative knowledge is represented as a network of small units of primitive knowledge called chunks. Figure 1 is a graphical display of a chunk encoding the addition fact that $3+4=7$ and some of its surrounding facts. These are some of the many facts that a child might have involving these numbers. Frequently, one encounters the question “What does it mean to understand 3 or to understand numbers in general?” The answer in ACT-R is quite definite on this matter: Understanding involves a large number of declarative chunks like those in Figure 1 plus a large number of procedural units which determine how this knowledge is used. According to the ACT-R theory, understanding requires nothing more or less than such a set of knowledge units. Understanding of a concept results when we have enough knowledge about the concept that we can flexibly solve significant problems involving the concept.

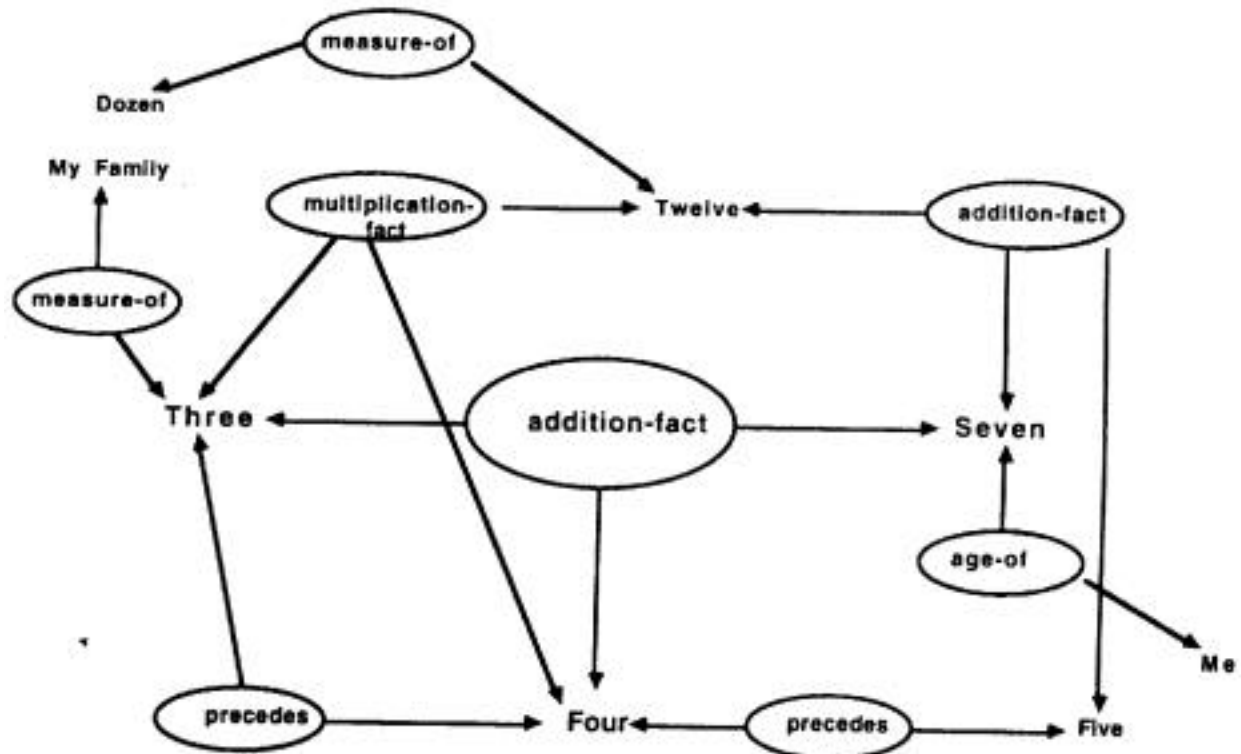


Figure 1: A graphical display of a chunk encoding the addition fact $3 + 4 = 7$.

Procedural knowledge, such as mathematical problem-solving skill, is represented by a large number of rule-like units called productions. Production rules are condition-action units which respond to various problem-solving conditions with specific cognitive actions. The steps of thought in a production system correspond to a sequence of such condition-action rules which execute or (in the terminology of production systems) fire. Production rules in ACT-R specify in their condition the existence of specific goals and often involve the creation of subgoals. For instance, suppose a child was at the point illustrated below in the solution of a multi-column addition problem:

$$\begin{array}{r} 534 \\ +248 \\ \hline 2 \end{array}$$

Focused on the tens column the following production rule might apply taken from the ACT-R simulation of multi-column addition in Anderson (1993):

IF the goal is to add n_1 and n_2 in a column
and $n_1 + n_2 = n_3$
THEN set as a subgoal to write n_3 in that column

This production rule specifies in its condition the goal of working on the tens column and involves a retrieval of a declarative chunk like the $3+4=7$ fact in Figure 1. In its action it creates a subgoal which might involve things like processing a carry. It is many procedural rules like this along with the chunks which in total produce what we recognize as competence in a domain like mathematics.

Goal Structures

It might seem that these chunks and productions are all separate, disjoint pieces of knowledge and that there is nothing in the ACT-R theory to produce the overall organization and structure in cognition. However, this ignores the contribution of the goal structure. Each task is decomposed into a sequence of subgoals which in turn may be decomposed into a sequence of subgoals. ACT-R maintains a stack of goals, onto which subgoals are added, and which are still remembered once the subgoals are achieved. Only the most recently added subgoal is used to select productions at any one point in time; once it is achieved it is removed from the goal stack. This hierarchical organization of subgoals and limited focus of processing imposes a strong order on the way in which knowledge is accessed and skills are applied. So, for instance, in multi-column addition there is goal structure that organizes the overall addition into specific column additions and processing carries. This produces an overall algorithmic-like process to solving multi-column addition. Of course, some tasks may have multiple possible goal structures and so permit for more variable behavior.

A simple example from education for which goal structures have played a prominent role is the case of multi-column subtraction. As it is typically taught in America, multi-column subtraction involves a subgoal of coordinating borrowing, especially from zero (Van Lehn, 1990). Many learning problems occur because these goal structures are not particularly obvious. Many of the bugs in multi-column subtraction are related to mastering the borrowing subgoal. For instance, when a child converts a 3 to a 13 but does not debit the next column, the child is failing to recognize the increment operator as part of the borrowing subgoal.

Learning Symbolic Structures

The important educational question concerns how these declarative and procedural units are learned. The ACT-R analysis of their acquisition is relatively straightforward. There are two ways in which declarative chunks can be acquired. The first way is encoding of information from the environment. For example, a child might encode the fact $3 + 4 = 7$ as part of reading an addition table. The second way is the storage of the results of past goals. For instance, at some point in time a child might have had the goal to find the sum of 3 and 4 and solved this by counting. The result of this counting process, the sum 7, would be stored with the goal chunk. Thus, the addition fact in Figure 1 could simply be a stored goal. This process of caching the results of past mental computations into chunks which can then be retrieved plays a major role in Logan's theory (Logan, 1988) of skill acquisition. He has accumulated a significant amount of data showing that it is important in the development of expertise.

Thus, with respect to acquisition of declarative knowledge, ACT-R holds that this can either be acquired in a passive, receptive mode (encoding from the environment) or an active, constructive mode (storing the result of past mental computations). The two modes of knowledge acquisition offer different advantages and disadvantages. Passive reception has the advantage of efficiency and accuracy. It is easier to read the sum of $3 + 4$ than to calculate it and there is not the danger of miscalculation. On the other hand, if one practices generating the knowledge, one is practicing a back-up strategy useful for when retrieval fails. However, according to ACT-R there is no inherent difference in the memorability of the two types of knowledge. There has been a fair amount of experimental work in memory on what is called the generation effect which is concerned with the supposed advantage of self-generated material (e.g., Burns, 1992; Hirshman & Bjork, 1988; Slamecka & Graf, 1978; Slamecka & Katsaiti, 1987). The generation effect is actually somewhat elusive and not always obtained. When it does occur, it seems related to redundancy of encoding. That is, generating knowledge for oneself is helpful only if the generation process produces multiple ways to retrieve the material. There are no magical properties conveyed upon a knowledge structure just because it was self-generated. If all things were equal it would be preferable to have children learn by generating the knowledge (due to the redundant encoding). However, because of difficulties of generation and dangers of misgeneration, things are not always equal and it can be preferable to tell the knowledge.

With respect to procedural knowledge, production rules are learned in ACT-R by a process we call **analogy**. For analogy to work in the ACT-R theory two things have to happen. First, the ACT-R must come upon a situation where it wants to solve a goal. Thus, in the case of the production rule above for addition the learner would come to a goal of wanting to perform multicolumn addition and be focused on adding two numbers in a column. Second, the learner needs an example of the solution of such a goal. So there might be an example of solving $4+5$ in some column. In this situation, the ACT-R analogy mechanism will try to abstract the principle in the example and form a production rule embodying this principle which can then be applied in the current situation. Once formed this production rule is then available to apply in other situations. Thus, ACT-R's theory of procedural learning claims that procedural skills are acquired by making references to past problem solutions while actively trying to solve new problems. Thus, it is both a theory of learning by doing and a theory of learning by example.

Simply providing the learner with examples is not sufficient to guarantee learning in the ACT-R theory. The sufficiency of the production rules acquired depends on the understanding of the

example. Example understanding can influence learning in two ways. First, it can influence which examples are retrieved for analogizing. When presented with a goal that cannot be solved with existing productions, ACT-R looks for previous examples that it has encountered involving similar goals. Obviously, the way it represents the previous examples and the current goal will affect which examples are retrieved. For instance, if the goal of solving one problem (e.g., solving algebra problems in class) is seen as very different from the goal of solving another problem (e.g., evaluating phone company rates), then the relevant example and accompanying solution procedure will not be retrieved.

Second, example understanding will influence the productions that are acquired by analogy to a given example. For instance, in the case of multi-column subtraction, one could understand an example involving a column subtraction of $8-3=5$ as either subtracting the top from the bottom number or as subtracting the smaller from the bigger number. The former understanding will produce the correct production rule (always subtract the top number from the bottom number), whereas the latter understanding will produce a buggy rule (always subtract the smaller from the larger). Similarly, Pirolli and Anderson (1985) show that students can learn very different rules for recursive programming from the same example programs.

Both of these factors place a premium upon the explanations that accompany examples in instruction. Chi, Bassok, Lewis, Reimann, and Glaser (1989) found that better learners of physics are those who more carefully study and try to understand examples. This “self-explanation” effect can be understood in terms of whether students generate adequate understandings of the examples.

Learning of both chunks and productions at the symbolic level in ACT-R are examples of all-or-none learning. In a single moment a new symbolic structure is formed and is permanently added to the system. However, simply having the symbolic knowledge there does not mean that it will be used successfully. The actual deployment of the knowledge depends on a set of activation processes which reflect the gradual accrual of strengths over experience. Anderson and Fincham (1994) and Anderson, Fincham, and Douglass (in press) document how rules gradually become effective with practice. In their task a rule needed to be practiced as often as 40 times before it became reliable. We will discuss activation processes next and then how underlying strengths are build up with experience.

The Activation Processes

The basic information-processing step in ACT-R is the firing of a production rule in which some declarative knowledge is retrieved and used to further the problem solution. The speed and success of this retrieval process depends on the level of activation of the chunks being retrieved and the strength of the production rules which are doing the retrieving. This determines the underlying fluency in performance. Fluency in display of knowledge is a critical educational goal for a number of reasons. First, in practical settings error-prone and slow performance is not acceptable. Second, development of more advanced competences requires fluency in the more basic competences. Third, students find fluent performance very reinforcing and are more willing to continue in the educational enterprise if they perceive themselves as successful in what they have tried to learn so far.

In ACT-R there is a mathematical theory of how activation and strength determine performance. These mathematics are important because they can be used to making precise predictions for

how practice results in increasing fluency. While our discussion does not depend on the exact equations in ACT-R, they are given below for completeness and we will qualitatively describe their implications. The probability of successfully retrieving a chunk is given by

$$\text{Probability} = \frac{1}{1 + Ce^{-c(A_i + S_p)}} \quad (\text{Accuracy Equation})$$

In this equation the critical quantity is $A_i + S_p$ where A_i is the level of activation of the chunk i being retrieved and S_p is the strength of the production p that retrieves the chunk. This equation makes accuracy in behavior an increasing function of this sum. As this function approaches 1 (perfect behavior) it becomes negatively accelerated such that the same increase in $A_i + S_p$ results in smaller and smaller performance increments.

The time to retrieve a chunk is given by

$$\text{Time} = Be^{-b(A_i + S_p)} \quad (\text{Latency Equation})$$

In this equation speed of performance is again a negatively accelerated function of the same quantity, $A_i + S_p$. As this sum increases, there are smaller and smaller improvements in speed. These performance equations imply that the sum, $A_i + S_p$, must reach a certain threshold before satisfactory performance is achieved. As we will see, both activation levels, A_i , and production strength, S_p , increase in a regular way with practice.

The activation of a chunk is related to both how well it has been learned and to how closely it is associated to the current context. For example, some chunks, such as your name, are so well learned that they have a very high activation regardless of context. By contrast, other chunks, such as the addition fact that $7+9=16$, will have a high activation only in the context of solving math problems involving 7 and 9. The influence of amount of previous learning is reflected in a quantity called **base-level activation**. The influence of associativeness to the current context is reflected in a quantity called the **associative activation**. Formally, the equation which specifies how these factors combine is:

$$A_i = B_i + \sum_j W_j S_{ji} \quad (\text{Activation Equation})$$

where B_i is the base-level activation (or strength) of the chunk i , the summation gives the associative activation, W_j in the sum is the attentional weighting for each of the elements j currently being focused upon, and S_{ji} is the strength of association from each element j . For instance, in the case of the addition fact in Figure 1, where we might be focused on the elements 3 and 4 in a column of a multi-column addition problem, 3 and 4 would be sources of activation

and S_{ji} would be the strengths of associations from these numbers to the target fact $3+4 = 7$. One might think from this equation that we could achieve any amount of activation by just focusing on enough sources and increasing the number of elements summed over in the equation. However, the evidence is that there are very strong bounds to the amount of contextual priming. This is represented in the ACT-R theory by constraining the W_j to sum to 1. Thus, in the example if 3 and 4 were the only sources of activation and they received equal weighting, they would both have a W_j of .5.

Thus, the performance of an ACT-R system turns on what determines the subsymbolic quantities of production strengths, S_p , base-level activations of chunks, B_i , and associative strengths S_{ji} . The next subsection will describe the learning processes which are responsible for determining these quantities.²

Subsymbolic Learning

ACT-R learns to concentrate its activation resources on those knowledge structures which experience indicates will to be useful. With respect to declarative structures, ACT-R gradually increases the base-level activation of a chunk as it is encountered more recently and frequently. The exact equation for base-level strength, which is based on a rich array of empirical results (Anderson & Schooler, 1991) is

$$B_i = \ln \sum_{j=1}^n t_j^{-d} \quad (\text{Base-Level Equation})$$

where the item has been encountered n times in the past at times t_1, t_2, \dots, t_n ago.³ Basically, each time the item is encountered there is an increment in its strength reflected in the summation. As time passes, these increments decay away--represented by negative power functions t_j^{-d} . The conglomerate of repetition (summation) and decay (negative power) is then passed through a logarithmic function to produce base-level activations. This is clearly a complex function, but when passed through the response functions given earlier (Accuracy Equation and Latency Equation) it leads to the prediction of three robust effects:

1. Power Law of Learning. As a particular skill is practiced there is a gradual and systematic improvement in performance which corresponds to a power law. This learning function is almost universally found. Figure 2 shows some data that we collected some time ago (Neves & Anderson, 1981) which illustrates power-law learning. We were looking at improvement in doing proofs as a function of the number of problems solved. As number of problems practiced increased there was a continuous decrease in time to do the next problem.

² Since the sum of the W_j is fixed they do not play a critical role in learning.

³ This complex formula turns out describe the log odds that a particular memory will be needed in the environment as a function of its past pattern of occurrences.

We have plotted the best-fitting power function to these data. Students started out taking over 20 minutes to do a problem and end up taking under 3 minutes. This is a speed up of almost an order of magnitude and reflects more than a quantitative change. The shift produces a qualitative change in how students approach mathematics problems because the reasoning skills become a practical part of their problem-solving repertoire and their attitudes change about their mathematical proficiency. Beck (described in Bruer, 1994) describes a similar qualitative change in student's self-confidence as their reading abilities become automatic through practice.

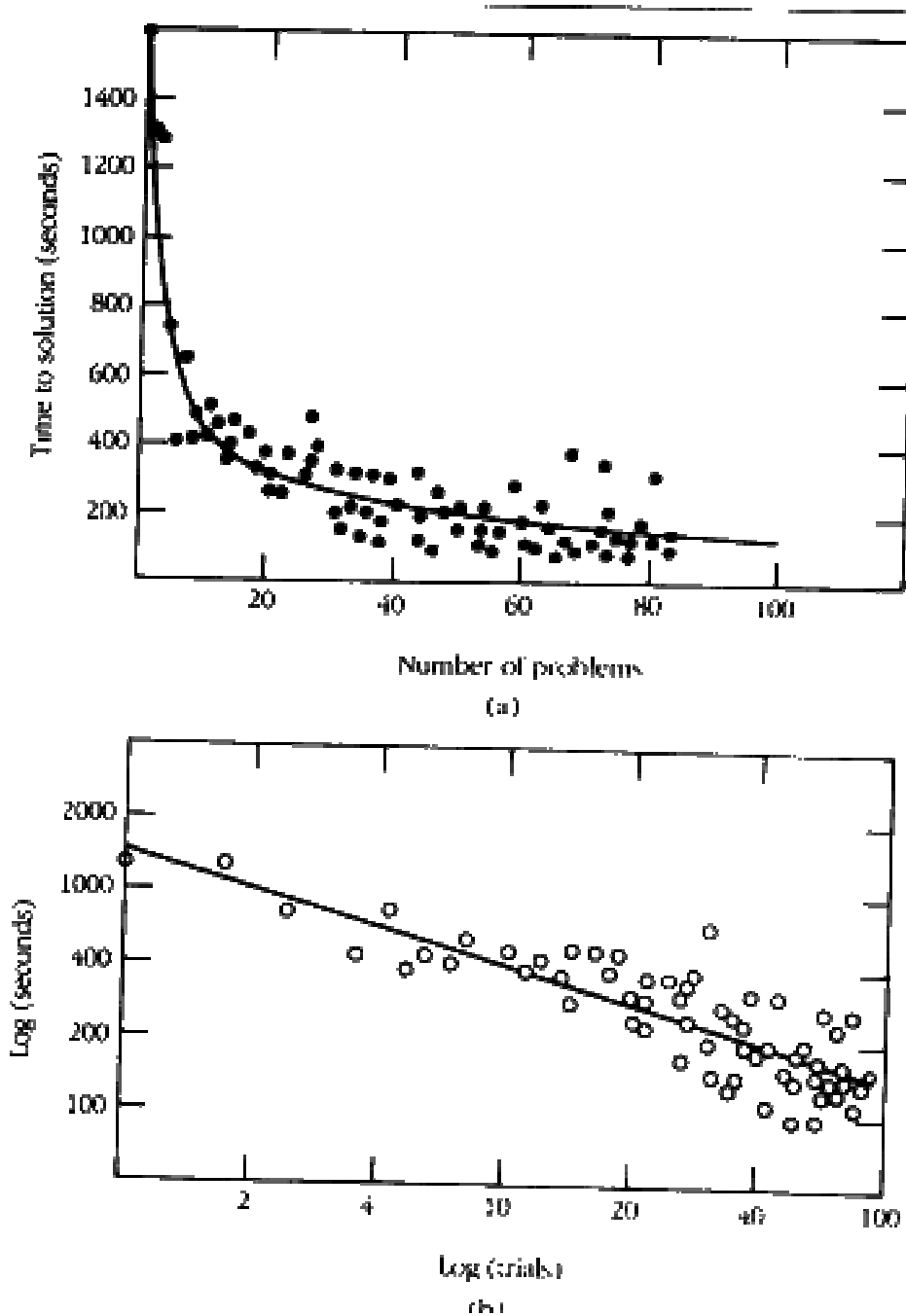


Figure 2: Effect of practice in performance of proof skills.

2. **Power Law of Forgetting.** As time passes performance degrades, also according to a power function. Again this effect is almost universally found (Rubin & Wenzel, in press). Figure 3 displays some data reported by Bahrick (1984) looking at the retention of Spanish over the years. The figure shows rapid initial loss followed by slow continuing loss. In some of his studies Bahrick finds at very long retention intervals a final rather steep drop off which he attributes to aging effects (personal communication).

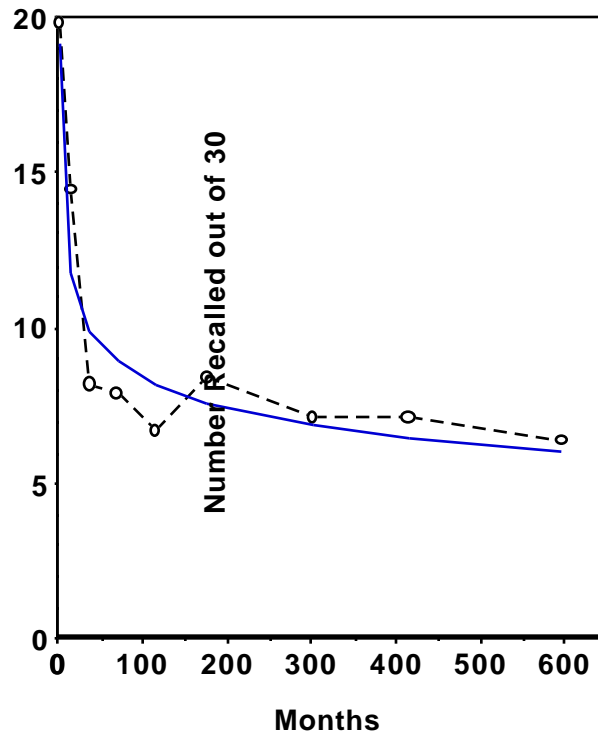


Figure 3: Bahrick’s data showing the combined impact of practice and retention interval on maintenance of foreign vocabulary.

3. **Multiplicative Effect of Practice and Retention.** Most important, the Base-Level Equation implies a relationship between the combined variables of amount of practice and duration over which the information must be maintained. This is illustrated in some other data from Bahrick and Hall (1991) which is displayed in Figure 4. Here they looked at retention of algebra material as a function of how much practice this material received. They looked at four groups--one who had minimal algebra in high school, another who have extensive high school experience, a third group who continued to use it in calculus in college, and a fourth group who used it in courses beyond calculus. As can be seen, the probability of retaining information is dramatically improved with the additional practice. Formally, the following equation captures the power law of practice, forgetting, and the multiplicative relationship implied by the ACT-R theory.

$$\text{Performance} = A N^c T^{-d} \quad (\text{Performance Equation})$$

This implies performance continuously improves with practice (N is amount of practice) and continuously degrades with retention interval (T is time). Most significantly the two factors multiply which means that increasing practice is a way to preserve the knowledge from the ravages of time. A large body of research is consistent with this performance equation (for a review see Anderson, 1995a).

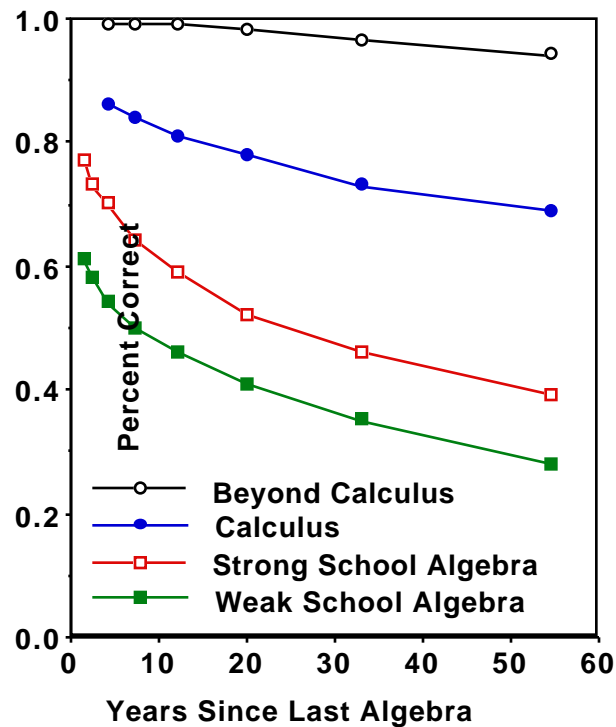


Figure 4: Bahrck & Hall’s data showing the impact of practice on retention of algebra.

For similar reasons, production strength, S_p , in ACT-R is related to its past usages by the same sort of equation as the Base-Level Equation:

$$S_p = \ln \prod_{j=1}^n t_j^{-d} \quad (\text{Production Strength Equation})$$

which implies again power-law learning, power-law forgetting, and a multiplicative relationship between practice and duration interval.

There is one other subsymbolic component determining activation levels. This concerns the strengths, S_{ji} , of association sources of activation j to declarative chunks i (see Activation Equation). This is what allows performance to be sensitive to the current goals and context. The exact learning process in ACT-R with respect to associative strengths is complex but basically it makes these strengths of associations be a function of the probability that a chunk will be needed when an element is in focus. However, there are limits to what we can get from contextual priming. A paradigm case of this is the acquisition of the addition facts. A particular number like 3 is associated with many facts and it can only provide limited priming for any particular one since the conditional probability of any fact given 3 is relatively low. The addition facts are notoriously difficult to learn. They can only be sufficiently mastered by enough practice to drive their base-level activation up. Moreover, when we next turn to examining the relationship between the ACT-R theory and educational goals, we will see that these S_{ji} 's are the least important of the subsymbolic activation parameters.

Strategy learning

Until now we have proceeded as if there were only one valid approach to solving any given problem. However, for most tasks, there are usually multiple methods or strategies for solving a problem. For example, Siegler and Jenkins (1989) found that when young children were asked to add two single-digit numbers, the children used four different strategies: guess an answer, retrieve the answer from memory, count up from the higher of the two numbers, or hold up as many fingers and count all of the held-up fingers. Most children in the Siegler and Jenkins study used all four strategies at least occasionally, and many children frequently switched the strategies they used from trial to trial. This high-degree of strategy variability has been found in a large number of domains. This raises the question: what determines strategy selection? Siegler and Shipley (1995) found that two factors were especially important in predicting strategy selection: the effort required to execute the strategy such that least effortful strategies were preferred, and the accuracy of the strategy such that the most accurate strategies were preferred. Moreover, these levels of accuracy and effort associated with different strategies change with practice. Thus, there will be changes with practice in which strategies tend to be preferred. For example, with practice in the domain, retrieval becomes less effortful and more accurate, and so children become more likely to rely on retrieval. Reder and Ritter (1992) also found that adults switch strategies from computation to retrieval as they practice two digit multiplication problems (e.g., 23×19). The importance of the strategy selection issue for education is that students will often rely on their own more primitive strategies to solve problems rather than use the taught strategy because the taught strategy is unpracticed, and therefore slow and error prone. For instance, MacLaren and Koedinger (1996) have shown that students will often prefer to use strategies like guess-and-check when solving word algebra problems rather than using formal algebra. Unfortunately, such backup strategies often do not transfer well to more complex problems, and the students are forced to either give up or then learn the competences they should have learned earlier.

In ACT-R, implementing strategy selection involves both a symbolic and a subsymbolic component. The symbolic component is very straightforward: different strategies are simply

represented with different productions (or sets of productions). For example, there is a production for retrieving from memory the sum of two numbers and there is a set of productions for counting out the sum of two numbers. When given a problem to solve, ACT-R selects among competing productions for different strategies and then executes the selected productions. The strategy selection process involves additional subsymbolic parameters. In particular, each production has two further subsymbolic parameters: expected effort and effected probability of success. The expected effort is the amount of effort that will be required to obtain a solution if the given production is selected. The expected success rate is the probability that a solution will be obtained if the given production is selected. These values are gradually learned through experience in using the productions (i.e., how much effort they required and how successful they were). ACT-R selects among productions by computing the tradeoff between success and effort and then selecting the production with the best tradeoff value. Using these features of ACT-R, Lovett and Anderson (1996) were able to provide a good fit to human strategy selection data. When we turn to educational consequences, we shall see that the effort and success features of strategy selection have similar educational implications as those for the strengths and activations of productions and chunks.

Summary

The ACT-R theory identifies two levels of analysis of knowledge, both of which will prove to be important to the goals of education. The first is the symbolic level where one acquires the knowledge structures that are the goals of an educational program. The second is the subsymbolic level where practice plays a critical role in building up base-level activations and production strengths. Only when these activations and strengths have reached adequate levels will the knowledge be successfully and rapidly deployed. As we have just reviewed, only when use of the knowledge has reached such robust levels will students come to consistently choose strategies which use this knowledge.

Educational Goals

Given the relatively simple character of ACT-R it might seem that it should be straightforward to apply it to education. The complication, however, is that it is not straightforward to specify what the goals of education are. The implications of the ACT-R theory depend very much on what one takes the goals of education to be. It is worthwhile contrasting two of the goals which the educational system seems to have. On one hand, if one were to take seriously the frequent public criticisms of the current state of education and its consequence for economic competitiveness and informed citizenship, one would take the goals of education to be to convey on students some general competences that will last them throughout their lifetime and be available over a wide range of contexts. One could reasonably argue that such enduring competences are the target of public education; short-term competences (e.g., specific knowledge relevant to playing a video game, using a locker combination, finding one's way in a city one is visiting, doing a task for a particular summer job) are not the targets of public education. The basic 3 R's are the classic examples of general competences that fit our definition. On the other hand, if one were to look at what students and their parents worry about with the approach of college and post-graduate schools, one would conclude the function of education is credentialing. That is, the goal is to do well on specific tests that occur at a specific time on well-specified subject domains.

Note this distinction between permanent competence and credentialing does not differ with the credentialing test. In particular our conclusions do not depend on whether the credentialing test is the infamous four-alternative forced-choice or a currently fashionable performance assessment.

The implications of the ACT-R theory are quite different depending on whether we take the goal of education to be to do well in life generally or do well on specific tests. If we set our focus on long-term competences that will be generally useful, the power laws of learning and forgetting, and the Performance Equation that embody them, become extremely important. By contrast much of what researchers in learning and memory study concerns the initial conditions for learning and the effects of various contexts on performance. However, the time scale desired for such competences makes largely irrelevant how well one does shortly after initial exposure to the material and make all important the amount and distribution of practice. The fact that the learning needs to be displayed broadly makes largely irrelevant any benefit one might get by exploiting special associations between sources and target memories. Thus, in ACT-R terms, factors controlling the strengths of associations, S_{ji} , between context j and chunk i , become less important because in general we cannot count on having specific elements j in the context.

On the other hand, if we took credentialing on some specific test as our goal, then the issues of initial acquisition and retrieval would become key and much more of the research from psychology would become relevant. This is because the learning does not have to last a lifetime nor be displayed across broad contexts. Good performance can be achieved with relatively little learning before that test if the initial conditions of learning are sufficiently favorable. Thus, things like mnemonic tests for learning foreign vocabulary can be very effective on a specific vocabulary test (Atkinson & Raugh, 1975). However, such mnemonics cease to be relevant if one practices a language enough to achieve true fluency. Moreover, without extensive practice there is evidence that mnemonics may result in worse long-term retention (Wang & Thomas, 1995).

Recently, the situated learning movement has attracted a lot of attention in education with demonstrations of how certain competences are specific to particular conditions of performance (Lave, 1988). Sometimes, these demonstrations are extended to argue that the general competence goal of education is impossible--that one cannot teach competences that are generally applicable. However, as Anderson, Simon and Reder (1996) document, this is certainly not the case. Rather, broad generality of application requires a great deal of practice in a broad range of situations. The situated demonstrations just illustrate the fact that sometimes it is easier to acquire a narrow competence than a broad competence.

So, we are left with a rather bland and perhaps distasteful recommendation that the general societal and utilitarian goals for education require practice and more practice. This conclusion is sanctioned by more than the ACT-R theory. The evidence is very strong that high degrees of competence only come through extensive practice. Bloom (1985a, b), Ericsson, Krampe, and Tesch-Römer (1993), Hayes (1985) and others have argued that high levels of proficiency in a domain are only achieved after at least 10 years of extensive practice. There is also a strong correlation across societies in the amount of time students devote to mathematics and the

competences of these students in mathematics (McKnight, Crosswhite, Dossey, Kifer, Swafford, Travers, & Cooney, 1990; Stevenson & Stigler, 1992; White, 1987). The differences can be striking. We have metered students in regular 9th grade algebra in the city of Pittsburgh (Koedinger, Anderson, Hadley, & Mark, 1995) and determined that they only average about 65 hours that year learning mathematics. This is a result of poor attendance patterns, failure to do homework, distractions in the classroom, etc. In contrast, in Japan students spend 175 class hours on mathematics in most of elementary school, 140 hours in junior high, and more hours again in high school (White, 1987). Stigler and Perry (1990) estimate that Japanese students are receiving instruction 90% of the official class time compared to only 46% of the time in American classrooms. Statistics like these add up to substantial differences in total learning time. It does not take a sophisticated theory to say that this is going to have significant consequences for learning. However, the ACT-R theory makes it clear that there is no magic bullet that allows some way out of these enormous differences in time on task. For competences to be displayed over a lifetime, time on task is by far and away the most significant factor.

There has been a long-standing strand of research in human memory looking at the advantage of mnemonics and various memory-enhancing strategies in terms of learning material. Such mnemonics strategies have been recommended for domains as far ranging as foreign vocabulary learning and learning of chemical formulas. However, the important thing to recognize is that these techniques speed the initial acquisition of the knowledge. Speed of the first steps on the learning curve becomes insignificant if one's goal is long-term possession of the knowledge. Such mnemonics drop out with practice and the critical factor becomes, not saving a relatively small amount of time in initial acquisition, but rather investing substantial amounts of time in subsequent practice. It is not clear that there is anything to be saved in subsequent practice by use of mnemonics.

Ericsson et al. have argued there is a distinction to be made between different sorts of practice with what they call "deliberate practice" being the type that leads to real learning gains. Deliberate practice is defined as involving motivated subjects, receiving informative feedback, with careful and continuous coaching and monitoring. In our view, these conditions simply guarantee that the subjects' time will be spent in learning the task rather than on other activities. For instance, if there is not careful monitoring the student can easily fall into practicing bad habits.

A general characterization of retention functions is that they reflect degree of original learning and are very insensitive to the conditions of original learning. One robust exception to this is the effect of spacing where it has been shown that better retention is obtained in conditions where the material is learned in a more spaced manner (see Anderson, 1995a for a review). This fact places the goal of credentialing in even greater conflict with the goal of permanent competence. Massing study before a test, which optimizes performance on the test, minimizes long-term retention. Also, there is evidence that broader transfer of training occurs if that training takes place over a wide range of contexts (Bjork & Richardson-Klaven, 1989) as would be predicted by the ACT-R theory of associative learning (the S_{ji} 's). Again the goal of credentialing for a specific test is in conflict with the conditions required to achieve general competence since

performance on a specific test is maximized by studying only in contexts similar to those of the testing situation.

We could take seriously the implications of ACT-R for credentialing tests and develop these into educational recommendations. In fact, most psychology researchers who have addressed educational implications of their research have been concerned with this goal thereby producing things like mnemonic strategies, methods for dealing with test anxiety, strategies for reading texts to take a test, etc. However, our characterizations of this as “credentialing” and our contrasting of this with the goals of “lifelong competence” signaled that we were just using it as a straw man to make a point. In what follows, we will be pursuing the implications of the ACT-R theory for the goal of enduring competences.

Task Analysis

The conclusion of the previous section is that practice (and nothing much else beyond spacing and varied practice) makes perfect. However, this leaves open the question of which skills the student is perfecting. Practice affects the subsymbolic encoding of the symbolic structures. It does not determine what these symbolic structures are. There is the real possibility that the student will not be practicing the symbolic structures that are the target of the instruction. Nothing in the ACT-R theory of activation and strength guarantees that, just because students put in a lot of time, they will be strengthening the right productions and chunks. As Ericsson et al. note, often practice does not make perfect when that practice fails to target the desired competence. In fact, if one is practicing the wrong competences, one might say “practice makes imperfect.”

In certain cases, communicating the right symbolic structures to the students is relatively unproblematic. Probably addition facts are a good case in point. All the student needs to do is put enough practice in (with access to the correct answers--otherwise the student might be practicing the wrong facts) and the right knowledge will become entrenched. However, in many cases there is not much feedback during the typical conditions of practice and the student might wind up entrenching the wrong knowledge structures. A well studied case of this is subtraction (Van Lehn, 1990; Young & O’Shea, 1981) where students can acquire the wrong rules and practice them to a state of perfection. An interesting alternative situation involves naive physics (Champagne, Klopfer, & Anderson, 1980; McCloskey, 1983) where students may have spent a life-time of practicing the wrong physics which is very hard to overcome when they come into the classroom. Often it is not a case of wrong knowledge but just failure to communicate any knowledge. We (Anderson, Boyle, & Yost, 1985; Anderson, Greeno, Kline, & Neves, 1981; Koedinger & Anderson, 1990) argued that a major problem in the acquisition of geometry proof skills was that it was hard for students to identify the component skills. Typically in geometry, students are shown complete proofs and left to figure out what problem solving steps underlie finding these proofs. A similar problem seems to haunt students’ attempts to master algebra word problems (Mayer, 1987; Singley, Anderson, Givens, & Hoffman, 1988).

Task analysis will often reveal prerequisite knowledge which is required in order for students to learn a new competence. Often this prerequisite knowledge has not been mastered by significant subsets of the population. An interesting example is knowledge of the number line and basic

operations on it. Case and Griffin (1990) found that many at-risk students lacked this knowledge which is a prerequisite to mastering early school mathematics. By explicitly teaching this aspect of mathematics to the students they dramatically increased their success at first-grade mathematics.

As knowledge domains become more advanced, their underlying cognitive structure tends to become more obscure. Thus, while it may remain easy to provide feedback on what the final answer is, it becomes difficult to provide feedback on the individual mental steps that lead to the final answer. Teachers often are unaware, at an explicit level, of what this knowledge is and do not know how to teach it to children. Some relatively basic skills suffer this problem. A good case in point is reading when one goes beyond the basic word identification skills. Palinscar and Brown (1984) were able to produce dramatic improvements in student's comprehension skills by introducing to the students and having the students practice the skills of summarizing, clarifying difficulties, question asking, etc. which are very valuable to engage in while reading. Apparently, these children had not been taught these skills and were unable to figure them out for themselves.

In other cases students have undoubtedly been "told" what they need to know but the telling did not take because of absence, inattention, and misunderstanding. In our work with students in high school mathematics we have seen students spend a year in an algebra class and maintain the belief that $3x+7$ means 3 times the quantity $x+7$. Frequently, it is because they were not there (physically or mentally) when the appropriate instruction was given. Other times it is because they never actively practiced this knowledge and soon forgot it.

There is nothing surprising in ACT-R about these results. For practice to be effective the right chunks and production rules need to be communicated. For this to happen, the teacher must know in some sense what these are and communicate them. Because of inevitable failures of communication, it is very helpful to monitor what the student is doing to assure that the communication has been successful, providing remedial instruction as necessary. Undoubtedly, one of the reasons for the success of human tutors and the occasional success of computer tutors (as we will describe in a later section) is that they provide this monitoring function.

This implies that there is a real value for an effort that takes a target domain, analyzes it into its underlying knowledge components, finds examples that utilize these components, communicates these components, and monitors their learning (for detailed examples see Anderson & Reiser, 1985; Anderson, Boyle, & Yost, 1985; Case & Griffin, 1990; Corbett, Anderson, & O'Brien, 1995; Koedinger, 1990). Unfortunately, cognitive task analysis receives relatively little institutional support. In psychology, there is little professional reward for such efforts beyond those concerned with basic reading and mathematics. The argument (which we have received from many a journal editor) is that such task analyses are studies of the characteristics of specific task domains and not of psychological interest. For experts in the various target domains (e.g. mathematics) the reward is for doing advanced work in that domain not for analyzing the cognitive structures underlying introductory skills in the domain. In education, such componential analyses have come to have a bad name based on the mistaken belief that it is not possible to identify the components of a complex skill. In part, this is a mistaken generalization from the failures of behaviorist efforts to analyze competences into a set of behavioral

objectives. Thus, we have a situation today where detailed cognitive analyses of various critical educational domains are largely being ignored by psychologists, domain experts, and educators.

Transfer, Insight, and Understanding

Much of the cognitive work applied to education flies under the banner of having students learn “with understanding” and contrasts itself with old “behaviorist” approaches of rote learning. Wrapped up in this dichotomy are two separate claims: 1) understanding comes in a rapid, transformational insight rather than through incremental learning; 2) understanding leads to more flexible application knowledge and greater transfer to novel settings. We will consider each of these claims in turn.

Are there rapid, transformational insights? There has been some experimental research on so-called “insight” problems (Metcalfe & Wiebe, 1987). One of the striking features of such problems is that subjects do not know they are close to producing a solution much before they actually produce a solution, suggesting that the solutions are obtained rapidly rather than incrementally. However, insight problems are defined as problems that require a single key insight for their solution. So it should not be surprising that it only takes a little time to formulate that one bit of knowledge. By contrast, non-insight problems (like doing a proof in geometry or solving an algebra problem) require developing multiple pieces of knowledge and students can judge when they have gotten some but not all of the problem.

In a careful analysis of the mutilated checkerboard, a famous insight problem, Kaplan and Simon (1990) studied the relationship of the critical insight to the rest of the problem solution. The mutilated checkerboard problem involves deciding whether it is possible to cover a mutilated checkerboard with dominos that each cover exactly two squares of the checkerboard. The checkerboard has two squares cut out from opposite corners. It turns out that it is impossible to cover the checkerboard because each domino covers a black and a white square and two opposite-corner squares of the same color have been removed. This is called the parity insight and subjects typically spend a lot of time trying unproductive paths before considering it. However, despite the apparent “insight” nature of this solution, it turns out that there are steps that lay the foundation for this insight such as choosing to consider the invariances in the problem. Moreover, when subjects did think of that insight, they still had to go through the process of working out its implications. Complete proofs of impossibility did not occur instantaneously but had to be developed given the decision to consider parity. So, even in so-called insight problems, bundles of knowledge do not come in one feel swoop but have to be worked out piece by piece.

If learning is not typically characterized by transformational insights, then what is the origin of this belief? The belief in moments of transformation in education is undoubtedly linked to the old belief in developmental psychology that children transit abruptly between stages. Siegler (1996) calls this the belief in the “immaculate transition”. Rather, as he documents with great care, development is always gradual and continuous. The same is true of education.

While understanding may not arise from transformational insights, one could still argue that understanding leads to more flexible application of knowledge and greater transfer to novel settings. Consider one of the classic contrasts of learning with understanding versus rote learning in Wertheimer’s (1945) comparison of teaching students to solve problems by rote or “insight”. Students given insight into the formula for the area of a parallelogram (by observing a construction) were able to transfer it to other, non-parallelogram figures for which the base times height formula is correct. Children just taught the formula were not able to transfer their knowledge.

To understand this result, consider what “understanding” might actually mean in an explicit theory like ACT-R. In ACT-R, understanding a concept means nothing more or less than having a rich network of highly available declarative chunks and production rules that can be used to solve problems involving those concepts flexibly in many contexts. The knowledge will transfer to new situations to the degree to which it is applicable to those situations. In the Wertheimer example, children in both conditions learned a set of facts and procedures. However, in the insight condition they were taught a different and richer set of facts which enabled the transfer. In fact, the insight instruction took longer reflecting the richer knowledge that was learned.

Singley and Anderson (1989) studied extensively the conditions under which knowledge learned in one kind of problem would transfer to solving another problem. They showed that transfer between domains was typically not all or none but varied with how much the two domains required use of the same knowledge. Understanding how knowledge will transfer between domains depends critically on cognitive task analyses where one examines the knowledge structures that the learner has acquired in one domain and assesses their applicability to another domain. One will get transfer to the extent there are shared cognitive elements.

In sum, understanding is best thought of as having a rich network of highly available declarative chunks and production rules that can be used to solve problems involving those concepts flexibly in many contexts. Each one of the knowledge elements has to be learned separately. Thus, understanding of a domain does not come in one fell swoop of insight but it rather built up bit by bit over time.

Cognitive Tutors

Our work with cognitive tutors (Anderson, Boyle, Corbett, & Lewis, 1990; Anderson, Corbett, Koedinger, & Pelletier, 1995) can be seen as reflecting an attempt to do cognitive task analyses and foster learning of the components according to the ACT-R theory. These projects begin with extensive task analyses in which we try to develop cognitive models to represent the competences that we are trying to teach. These cognitive models take the form of computer simulations (ACT-R-like--although most of our cognitive models have not actually been built in ACT-R but in somewhat simplified simulation languages) which are capable of solving the classes of problems that we ask students to solve. We develop instruction to communicate the declarative and procedural knowledge that has been identified. Our tutors then monitor students’ problem-solving and try to diagnose what the students know and do not know, providing help and scaffolding to deal with their weaknesses and dynamic instruction to repair the holes in their knowledge. These cognitive tutors gain their power, first, by their analysis of the underlying knowledge which allows instruction to be directed at what needs to be learned and, second, by their ability to monitor the student's problem solving and to make effective diagnosis. As a consequence of these features the tutors can insure that the students’ learning time is profitably invested. In Ericsson et al.’s terms, they help guarantee that the student is engaged in “deliberate practice”.

Our cognitive tutors reflect the two fundamental insights of the ACT-R theory for instruction -- the importance of task analysis to identify the right knowledge structures and the importance of practice to permanently establish those knowledge structures. We have accumulated a great deal of evidence over years of research with our tutors for these aspects of the ACT-R approach. This evidence is fundamentally of two kinds. First, we have been able to show that the regularity of

learning behavior in our tutors reflects what would be expected from the ACT-R approach. Second, we have been able to show that these tutors produce significant achievement gains. We will describe examples of both types of evidence.

Figure 5 shows some recent data from Corbett, Anderson, and O'Brien (1995) from the LISP tutor. Subjects are asked to write a number of LISP functions and the figure shows their error rate at each point in each function across the early material. Subjects show a lot of jumps and drops in their error rates at different points. This behavior appears at first glance to be totally erratic and random. Most importantly, there appears to be very little learning over the course of practice. The factor underlying this apparent lack of improvement is that the exercises are becoming progressively more difficult over time and are introducing novel elements. When one applies a task analysis to these exercises, breaking each task into the productions required for solution, the error patterns become quite predictable. One can track learning of individual production rules. One rule may get introduced early but not be practiced again until much later. Another rule might not be introduced until the middle of the sequence but then receive a number of opportunities for practice. One can look at the error rate the first time a production needs to be used, the second time, etc. Figure 6 shows the pattern that results when errors are plotted as a function of amount of practice with specific rules. The probability of error at each point in the solution is a simple power function of the number of times that production has been practiced in an example in the past.

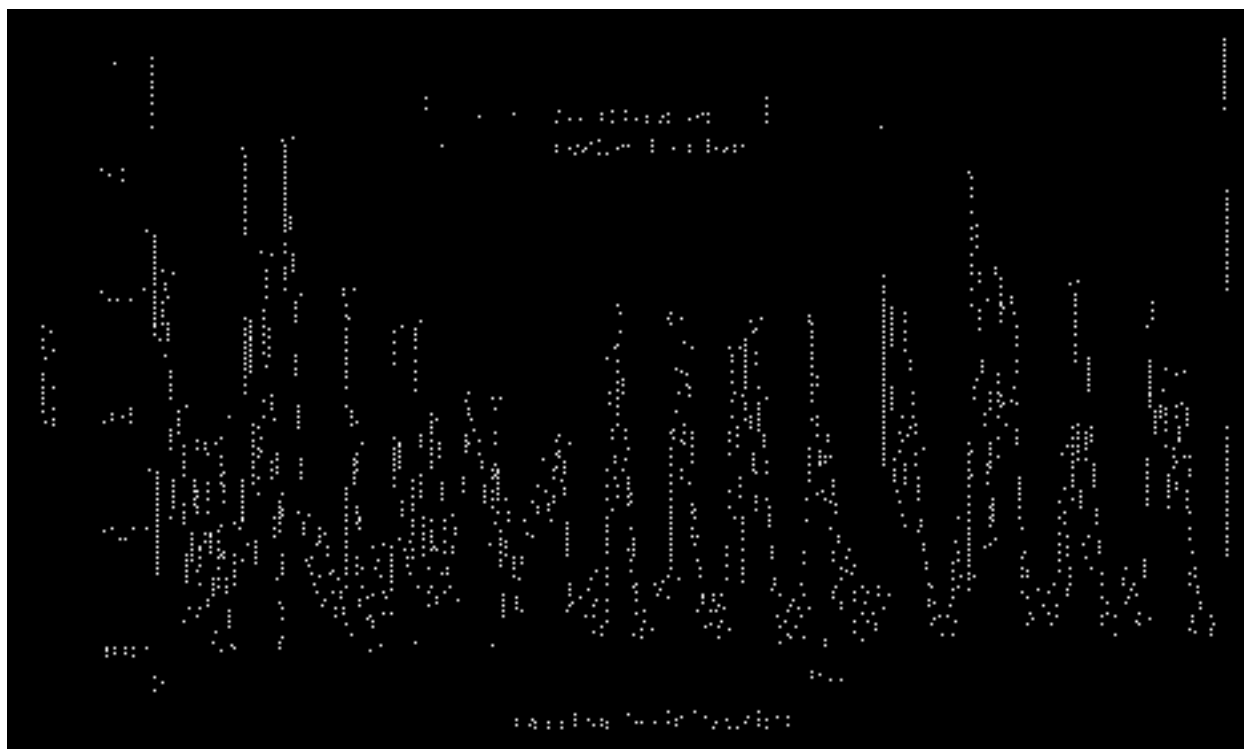


Figure 5: Data from Corbett, Anderson, and O'Brien illustrating the ability of the ACT-R theory to trace the complex trajectory of learning.

Using this cognitive task analysis and production-based learning function, the complex error pattern in Figure 5 becomes perfectly predictable. The success in accounting for complex error profiles like Figure 5 is for us the most compelling support for a componential analysis of a

complex competence. This is one example of many successes (see Anderson, 1993) that we have had at task decomposition. We are able to show that we can take the learning of a complex task and decompose it into a number of simple learning curves. It has gotten to the point where we now use these learning curves as a design mechanism. That is, if we deploy a tutor and it does not yield smooth learning curves we take this as evidence that we have not achieved the appropriate knowledge decomposition and we do a reanalysis of the domain to come up with the right units of knowledge that will yield smooth learning curves.

The educational effectiveness of our tutors is very much a function of the accuracy of the task analysis, the quality of the diagnosis and feedback, and whether they are appropriately deployed in the classroom (see Anderson, Corbett, Koedinger, & Pelletier, 1995 for a review). In tests of our well-tuned LISP tutor in the laboratory we have shown that students can achieve the same level of competence in one-third the time as traditional education. When we deploy our tutors in real classrooms typically we compare students with the same amount of instruction and just manipulate access to the tutor. Here we have gotten about 1 standard deviation or one letter grade gains with geometry tutors, algebra tutors, and programming tutors. So, as a general summary, our tutors are usually quite effective in providing achievement gains. Just how effective they are depends on how well they are designed and how well they are deployed.

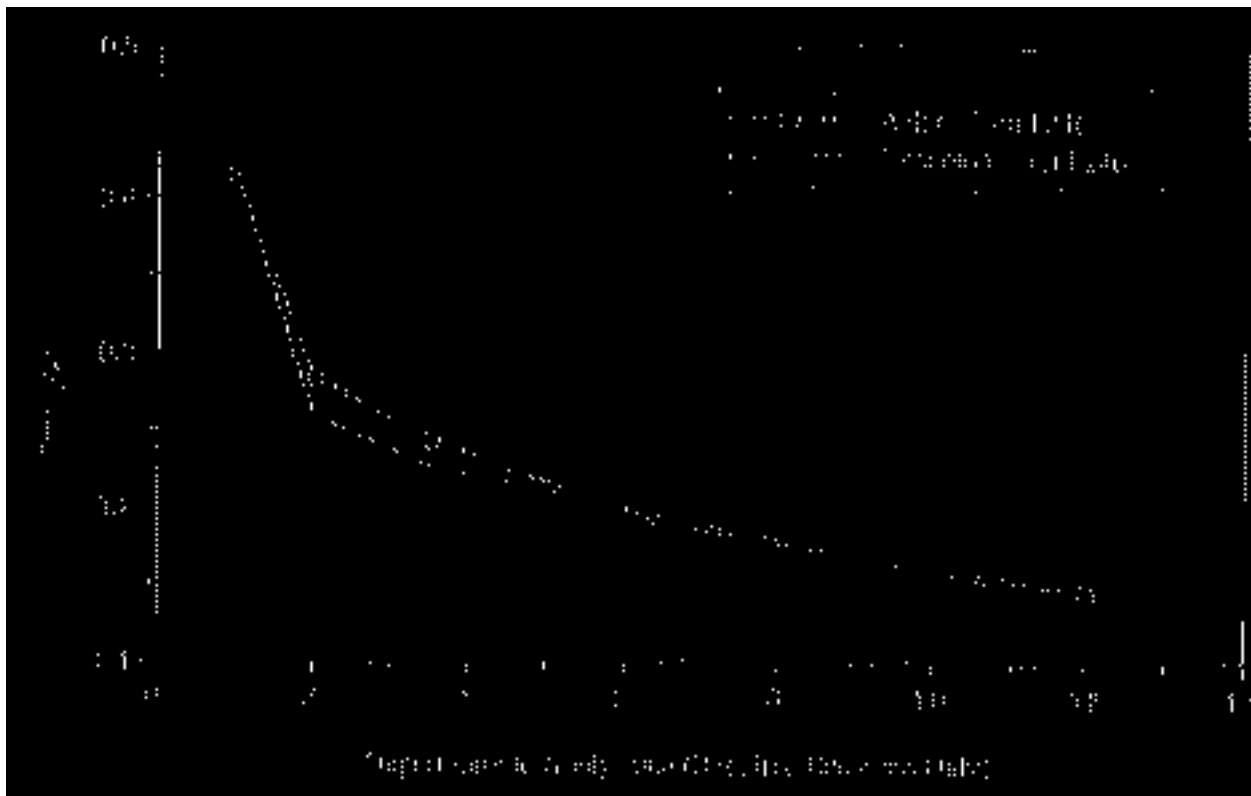


Figure 6: The average learning curves underlying the performance functions in Figure 5.

Our tutors are effective to the extent that they make efficient use of the student's learning time. The critical variable is how much practice students give to underlying components of knowledge. Analysis of what students actually do during study and homework time in non-tutored situations

reveals that often two-thirds of the time that they devote to studying is not spent practicing the critical components. This may not be any fault of the students who may be trying very hard. However, much of this time is being wasted recovering from temporary error states and misconceptions. Our tutors are effective substantially to the extent that they eliminate wasted time. In addition they insure that each desired competence has reached sufficient performance levels so that no targeted skills “fall between the cracks”.

Three components in our tutor design are key in determining their effectiveness at managing student’s learning time so that it stays profitable. First, as we have already mentioned, there is the accuracy of the underlying cognitive model and significant achievement gains can be gotten by improving it. Second, there is the issue of how well the instruction communicates to the student the contents of the target model. When the student makes an error and cannot produce the correct action, the tutor must suggest to the student what action should be taken in such a way that the desired production is learned. Along these lines, McKendree (1990) was able to get a substantial improvement in the geometry tutor by improving the help messages and the error messages. Third, there is the issue of how well the model does at diagnosing where the student is in the problem solving and what the student does and does not know. Basically, all three of these factors are manifestations of the fundamental issue of how well we can get the student and tutor aligned. By no means do we think we have achieved the best possible solutions to these alignment issues and they remain a current focus of research. Below we elaborate a bit more on each of these factors.

The development of an accurate cognitive task analysis has been and continues to be the most time intensive part of developing instruction. This is because we must examine each domain we want to teach anew and cannot carry over knowledge from one domain to another. This is not a unique problem of the tutoring approach. Any attempt to do instruction informed by cognitive models is going to face this investment. As we lamented earlier, there does not seem to be any professional organization to foster development of such cognitive task analyses.

The issue of improving help and error messages has received the least systematic analysis. In large part, this reflects the fact that its systematic treatment would require a theory of natural language processing. This would create a subgoal larger than the rest of the cognitive modeling enterprise. Our ability to make progress in the tutoring analysis has rested on our decision to take intuitive solutions to the communication task. However, this undoubtedly means that our tutors are not as effective as they might be.

ACT-R does advocate an emphasis on the use of examples for knowledge communication and much of the instruction that accompanies our tutors is based around specific examples. We accompany these examples with instruction to highlight the critical information. This instruction is not that different than the help messages that the tutors offer about critical aspects of problems on which a student is working.

It is in the diagnosis component of our tutors that we think we require more of a cognitive theory than simply the recommendations of task analysis and practice. If we want to individualize the instructional experience to fit a particular student we must try to examine in detail the knowledge state of that student. The ACT-R theory implies precise ways to use the speed and accuracy of a

student's responses to diagnose the relative strengths of the knowledge units. We can use this diagnosis of the relative strengths of knowledge units to assess the relative plausibility of different interpretations of students' actions. For example, consider the following algebra problem:

$$2x=6$$

Suppose the student responds with $x=4$. In this case, there are at least two possible interpretations to the cause of this error. First, student may have erroneously decided to subtract 2 from both sides to get 4. Second, the student have correctly decided to divide 2 from both sides but then may have erroneously retrieved the answer 4 to the division problem of $6/2$. The prior history of the student can allow one to judge how well the student has mastered the relevant algebra procedure versus the relevant division fact. One can use this diagnosis to provide appropriate instructional feedback.

Such diagnosis requires that we pay attention to the history of learning and to the details of the current problem. For instance, in the example above, if the student has a history of problems with their 2 times table, then $6/2=4$ becomes the more plausible interpretation of the error. It is for this diagnosis task that much of the research on learning and memory becomes relevant (for example, priming from recent problems, similarity to past problems, mnemonics used to learn arithmetic facts). For instance, if 4 has been primed as a recent answer to a problem, the $6/2=4$ interpretation also becomes more plausible (Campbell, 1991). Thus, such cognitive psychology research may not be relevant directly to the goal of teaching enduring competences where targeted practice is all that is key, but it is relevant to the subgoal of diagnosis in service of targeting the practice.

There is substantial potential for using complex cognitive theories to guide diagnosis. The real issue is to find behavioral indicants which will allow us to use these theories. For instance, we have recently been exploring the prospect of using eye-movement data to inform cognitive modeling. By recording where students are looking and how long they are looking, we have fine-grained evidence about what they are thinking. We can then bring to bear theories about language processing, visual reasoning, and other aspects of cognitive processing to the task of inferring what the student knows. There is the very real potential that such information will allow the cognitive tutors to be more sensitive to the individual student than human tutors.

Summary

The implications of the ACT-R theory for educational practice are very much like its implications for the learning of students: There is no substitute for a lot of careful work. It is important to analyze the competences that one wants to teach to the students and carefully monitor their learning to make sure that the right skills are being acquired. The ACT-R theory of performance and learning sanctions this basic approach and places an emphasis on practice. But it places little emphasis on the conditions of learning, except for the importance of accurate diagnosis and feedback. The details of the theory become especially relevant only in the actual

monitoring of the competence and diagnosis of what the student knows. This can be an important role, but it is important because it helps assure the practice is well spent.

Much of what one finds in educational research, which often comes with claims of a basis in cognitive psychology, is orthogonal (not necessarily opposite) to these recommendations. To take the area with which we are familiar with, mathematics education, the majority of the effort in the field appears directed towards change of the curriculum and methods of assessment. Much of this curriculum reform effort is justified by the observation that technology has changed what mathematical competences are needed. However, changing the curriculum does not change the value for componential analysis nor the need of students to master the components. One does read claims that such cognitive analyses are no longer relevant to the new curriculum standards but such claims are simply false. Instead, we have found that the learning of the new curriculum consists of componential learning just like the old curriculum (Koedinger, Anderson, Hadley & Mark, 1995). It is every bit as important to practice these components of “new look” curriculum to achieve a high level of mastery as it was to practice the components of the “traditional” curriculum.

In closing, one might ask what the implications of our tutoring work has been for the ACT-R theory. By far and away the most important outcome has been confirmation of the knowledge decomposition assumption in ACT-R—that one can take a complex domain and analyze it into productions and chunks. The tutoring work has had relatively little impact on the detailed assumptions in ACT-R about such things as activation computations and the subsymbolic level of strategy selection. This may be related to the fact that computational limitations have prevented us from simulating the subsymbolic levels in our tutors. Recent improvements in the efficiency of ACT-R’s implementation and ever-faster computational hardware may enable us to use this level of ACT-R in future tutors. As described above, the major function of such detailed simulations would be to enable more accurate cognitive diagnosis. However, such improved computational power from our tutors will not eliminate the need for the extensive task analysis on the part of the educators nor the extensive practice on the part of the students.

References

- Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1995a). *Learning and Memory*. New York: Wiley.
- Anderson, J. R. (1995b). *Cognitive Psychology and Its Implications: Fourth Edition*. New York: Freeman.
- Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355-365.
- Anderson, J.R., Boyle, C.F., & Yost, G. (1985). The geometry tutor. *Proceedings of IJCAI*, 1-7.
- Anderson, J. R., Boyle, C. F., Corbett, A., Lewis, M. W. (1990) Cognitive modeling and intelligent tutoring. *Artificial Intelligence*, *42*, 7-49.
- Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of Learning Sciences*, *4*, 167-207.
- Anderson, J. R. & Fincham, J. M. (1994). Acquisition of Procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1322-1340.
- Anderson, J. R., Fincham, J. M., & Douglass, S. (in press). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Anderson, J.R., Greeno, J.G., Kline, P.K., & Neves, D.M. (1981). Acquisition of problem solving skill. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Erlbaum.
- Anderson, J.R., & Reiser, B.J. (1985). The LISP tutor. *Byte*, *10*, 159-175.
- Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396-408.
- Anderson, J. R., Simon, H. A. & Reder, L. M. (1996). Situated learning and education. *Educational Researcher*, *25*, 5-11.
- Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, *104*, 126-133.
- Bahrack, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, *113*, 1-24.
- Bahrack, H. P. & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General*, *120*, 20-33.
- Bjork, R. A. & Richardson-Klavehn, A. (1989). On the puzzling relationship between environment context and human memory. In C. Izawa (Ed.) *Current issues in cognitive processes: The Tulane Flowerree Symposium on Cognition*. Hillsdale, NJ: Erlbaum.
- Bloom, B. S. (1985a). (Ed.) *Developing talent in young people*. New York: Ballantine Books.
- Bloom, B. S. (1985b). Generalizations about talent development. In B. S. Bloom (Ed.), *Developing talent in your people*, (pp. 507-549). New York: Ballantine Books.
- Bruer, J. T. (1993). *Schools for thought: A science of learning in the classroom*. Cambridge, MA: MIT Press.
- Burns, D. J. (1992). The consequences of generation. *Journal of Memory and language*, *31*, 615-633.
- Case, R. & Griffin, S. (1990). Child cognitive development: The role of control conceptual structures in the development of scientific thought. In C.A. Hauert, (ed.) *Developmental*

- Psychology: Cognitive, Perceptuo-Motor, and Neurophysiological Perspectives*. North Holland.
- Campbell, J. I. D. (1991). Conditions of error priming in number-fact retrieval. *Memory & Cognition*, *19*, 197-209.
- Champagne, A. B., Klopfer, L., E., & Anderson, J. H. (1980). Factors influencing the learning of classical mechanics. *American Journal of Physics*, *48*, 1074-1079.
- Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182.
- Corbett, A.T. and Anderson, J.R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*, 253-278.
- Corbett, A.T. and Anderson, J.R. and O'Brien, A.T. (1995). Student modeling in the ACT Programming Tutor. In P. Nichols, S. Chipman and B. Brennan (eds.) *Cognitively Diagnostic Assessment*. (19-41). Hillsdale, NJ: Erlbaum.
- Ericsson, K. A. Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363-406.
- Hayes, H. R. (1985). Three problems in teaching general skills. In J. Segal, S. Chipman, & R. Glaser (Eds.) *Thinking and learning* (Vol. 2). Hillsdale, NJ: Erlbaum.
- Hirshman, D. L. & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 484-494.
- Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*, 374-419.
- Koedinger, K. R. (1990). Tutoring concepts, percepts, and rules in geometry problem solving. Unpublished dissertation. Department of psychology, Carnegie Mellon University, Pittsburgh, PA.
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, *14*, 511-550.
- Koedinger, K. R., Anderson, J.R., Hadley, W.H., & Mark, M. A. (1995). Intelligent tutoring goes to school in the big city. In Proceedings of the 7th World Conference on Artificial Intelligence in Education, (pp. 421-428). Charlottesville, VA: Association for the Advancement of Computing in Education.
- Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527.
- Lovett, M. C. & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology*, *31*, 168-217.
- Mayer, R. E. (1987). *Educational psychology: A cognitive approach*. Boston: Little, Brown.
- MacLaren, B. A. & Koedinger, K. R. (1996). Toward a dynamic model of early algebra acquisition. Paper presented at the 1996 European Artificial Intelligence in Education conference.
- McCloskey, M. (1983). Intuitive physics. *Scientific American*, *248*, 122-130.
- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*, 107-157.
- McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human-Computer Interaction*, *5*, 381-414.

- McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J. & Cooney, T. J., (1990) *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. Champaign, IL: Stipes Publishing Co.
- Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and non-insight problem solving. *Memory & Cognition*, *15*, 238-246.
- Neves, D.M., & Anderson, J.R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Erlbaum.
- Palinscar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117-175.
- Pirolli, P.L., & Anderson, J.R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 136-153.
- Reder, L.M. & Ritter, F. (1992) What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435-451.
- Rubin, D. C. & Wenzel, A. E. (in press). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
- Siegler, R. S., & Jenkins, E. A. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.
- Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modeling*, (pp. 31-76). New York: Academic Press.
- Singley, M. K. & Anderson, J. R. (1989). *Transfer of Cognitive Skill*. Cambridge, MA: Harvard University Press.
- Singley, M. K., Anderson, J. R., Gevins, J. S., & Hoffman, D. (1989). The algebra word problem tutor, *Artificial Intelligence and Education*, 267-275.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *4*, 592-604.
- Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language*, *26*, 589-607.
- Stevenson, H. W. & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York: Summit Books.
- Stigler, J. W. & Perry (1990). Mathematics Learning in Japanese, Chinese, and American Classrooms. In J. W. Stigler, R. A. Shweder, and G. Herdt (Eds.) *Cultural Psychology*. New York: Cambridge University Press.
- Van Lehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.
- Wang, A. Y., & Thomas, M. H. (1995). The effect of keywords on long-term retention: Help or hindrance? *Journal of Educational Psychology*, *87*, 468-475.
- Wertheimer, M. (1945). *Productive thinking*. New York: Harper & Row.
- White, M. (1987). *The Japanese education challenge: A commitment to children*. New York: Free Press.
- Young, R. & O'Shea, T. (1981). Errors in children's subtraction. *Cognitive Science*, *5*, 153-177.