Vandenberg, S. G. (1971). What do we know today about the inheritance of intelligence and how do we know it? In R. Cancro (Ed.), *Intelligence: Genetic and environmental influences.* New York: Grune and Stratton.
Vernon, P. E. (1972a). The usefulness of "creativity tests." *The School Guidance Worker,* 27:30–35.
Vernon, P. E. (1972b). The validity of divergent thinking tests. *The Alberta Journal of Educational Research,* 18:249–258.
Walsh, D. A. (1986). Aging and human visual information processing. *Geriatric Opthamology,* 2:29–35.
Wechsler, D. (1941). *The measurement of adult intelligence,* 2d ed., Baltimore: Williams and Wilkins.
Witkin, H. A., and Goodenough, D. R. (1981). *Cognitive studies: Essence and origins.* New York: International Universities Press.
Wolman, B. B. (Ed.) (1985). *Handbook of intelligence.* New York: Wiley.
Wood, W. E., and Elias, M. E. (Eds.) (1982). *Alcoholism and aging: Advances in research.* Boca Raton: CRC Press.
Woodcock, R. W. (1978). *Development and standardization of the Woodcock-Johnson psycho-educational battery.* Allen, Tx.: DLM Teaching Resources.
Woodcock, R. W., and Johnson, M. B. (1977). *Woodcock-Johnson psycho-educational battery.* Allen, Tx.: DLM Teaching Resources Corp.
Woodcock, R. W. (1989) Woodcock-Johnson psycho-educational battery—revised.

Kyllonen, P. C. & Shute, V. J. (1989). A taxonomy of learning skills. In P. L. Ackerman, R. J. Sternberg & R. Glaser (Eds.), Learning and Individual Differences (pp. 117-163). New York, NY: W.H. Freeman

4

A Taxonomy of Learning Skills

PATRICK C. KYLLONEN[1] *Institute for Behavioral Research, University of Georgia*

VALERIE J. SHUTE *Air Force Human Resources Laboratory*

What is the relationship between intelligence and learning ability? This question engaged contributors to the 1965 conference on *learning and individual differences,* and, we believe, the sophistication of the answer to this question, perhaps as clearly as to any other, highlights exactly how far our theories have come over the last twenty years.

Certainly the prevalent position among the contributors to the 1965 conference, and indeed the general opinion until recently, was that there is no relationship between intelligence and the ability to learn or, perhaps, that the relationship is weak at best. This position reflects conclusions drawn from the widely cited series of studies by Woodrow (1946), who found that with extended practice on a variety of learning tests (such as canceling tasks, analogies, addition), the performance of brighter students did not improve at a rate substantially greater than that shown by poorer students. Woodrow's studies are no longer viewed as incontrovertible in addressing the intelligence-learning

[1]Currently at Air Force Human Resources Laboratory, Brooks Air Force Base, Texas.

issue, primarily because of problems with the measures of learning ability he employed: his learning tasks may have been too simple (Humphreys, 1979; Campione, Brown, and Bryant, 1985), and his conception of learning as improvement due to practice was too simplistic. Had he selected other kinds of learning tasks and measured learning with other performance indices, his results might have been quite different, as subsequent investigation has shown (Snow, Kyllonen, and Marshalek, 1984).

We may draw a general conclusion here: to address questions regarding learning ability, such as the question of its correlates and its dimensionality, it is important to have a clear idea of exactly what is meant by learning ability to the extent that one can specify learning indicators. Problems and confusions such as those introduced by Woodrow could have been resolved by selecting learning indicators from an agreed-upon taxonomy of learning skills. To clarify this point, for the purposes of this paper we distinguish learning *abilities* from learning *skills*. We define *abilities* as individual-difference dimensions in a factor analysis of learning tasks. We define *skills* as candidate individual-difference dimensions which are presently only conceptually distinct. In this way, we believe that proposing learning skills logically precedes establishing the individual differences dimensions underlying learning. Proposing a taxonomy of learning skills should assist in determining the dimensions of learning ability. (We realize that our use of the terms *abilities* and *skills* may be somewhat idiosyncratic.)

There are many potential benefits to having a widely accepted taxonomy of learning skills. Consider Bloom's *Taxonomy of Educational Objectives* (1956). Its primary purpose was to serve as an aid, especially to teachers, for considering a wider range of potential instructional goals and for considering means of evaluating student achievement consistent with those goals. Although the taxonomy has been criticized for vagueness (what exactly is analysis anyway?) (Ennis, 1986), it has served teachers well over the last thirty years, at least as demonstrated by its continued inclusion in teacher training curricula. Its main effect has probably been to encourage instructing and testing of higher-order thinking skills (analysis, synthesis, evaluation). A taxonomy of learning skills could have a parallel effect in encouraging the development of instructional objectives concerned with teaching higher-order learning skills.

Fleishman and Quaintance (1984) have outlined a number of ways, both scientifically and practically, in which a performance taxonomy in psychology would be beneficial. The main scientific benefit would be that results from different studies using different methods could more easily be compared and synthesized. Study A finds that some manipulation drastically affects performance on task X whereas study B finds that the same manipulation has no effect on performance of task Y. Are the studies contradictory or compatible? A taxonomy could help decide.

The main practical benefit of having a taxonomy of learning skills is that consumers of research findings could more easily determine the limits of generalizability from current research findings to an immediate practical problem. For example, it would be convenient to be able to produce learnability metrics for any kind of learning task, either in the classroom (for example, a particular algebra curriculum) or outside the classroom (such as a new word-processing system). A taxonomy of learning skills would be an important first step toward achieving a generally useful learnability metric system.

There are also more specific motivations for the immediate development of a taxonomy of learning skills. The National Assessment of Educational Progress (NAEP, "The Nation's Report Card") is a biennial survey of student achievement in areas such as mathematics, science, and computer science designed to provide information to Congress, school officials, and other policy makers regarding the state of American education. In recent years there has been increasing attention given to the assessment of higher-order skills in these subject matters (Murnane and Raizen, 1988). It is likely that because of political pressures this effort will continue with or without a taxonomy, but a taxonomy of learning skills could assist in the development of new, more refined test items to measure learning skills relevant to math and science.

Perhaps the most conspicuous benefits of having a viable taxonomy of learning skills, however, would be realized in the burgeoning domain of intelligent computerized tutoring systems (ITS). A number of such systems have been developed (Yazdani, 1986), and the potential for generalizing and synthesizing results across the different systems is being seen as increasingly critical (Soloway and Littman, 1986). Too often, researchers caught up in the excitement of developing powerful, innovative instructional systems have neither the interest nor the expertise for systematically evaluating those systems. There have been a few small-scale evaluation studies of global outcomes (Anderson, Boyle, and Reiser, 1985), but the field could obviously benefit from an accepted taxonomy. System developers could state what kinds of learning skills were being developed, and evaluators could determine the degree of success achieved. In this way, a taxonomy could provide a useful metric by which to compare and evaluate tutors as to their relative effectiveness not only in teaching the stipulated subject matter, but also in promoting more general learning skills.

Intelligent tutoring systems would benefit from a learning taxonomy in a second way. Because of the precision with which instructional objectives may be stated, the degree of tutorial control over how these objectives guide instructional decisions, and the precision with which student learning may be assessed, the ITS environment enables the examination of issues on the nature of learning that investigators simply were unable to address in the past. Educational research has been notoriously plagued with noisy data due to the very nature of field research and the inherent lack of control over the way instructional treatments are administered and learning outcomes measured. The controlled ITS environment thus offers new promise as the ideal testbed for evaluating fundamental issues in learning. With these systems we now have the capability of generating rich descriptions of an individual learner's progress during instruction. A taxonomy should help in determining exactly what indicators of learning progress and learner status we ought to be producing and examining. The acid test of the utility of any learning taxonomy is whether it could actually be used to assist in such an endeavor. The goal of this chapter is to propose such a taxonomy. We begin by looking at what has been done thus far.

A TAXONOMY OF LEARNING TAXONOMIES

Investigators have adopted various approaches to the development of learning taxonomies. One way of organizing these approaches, which we will apply here, is by the categories of (a) *rational*, based on a conditions-of-learning analysis, (b) *correlational*, based on an individual-differences analysis, and (c) *model-based*, from formal computer simulations of learning processes.

Rational Taxonomies

Rational taxonomies are by far the most common. Examples of this type are taxonomies proposed by Bloom (1956), Gagné (1985), Jensen (1967), and Melton (1964). Proposed taxonomies are based on a speculative, rational analysis of the domain, and frequently the analysis applied is of a conditions-of-learning nature. That is, the proposer defines task categories in terms of characteristics that will foster or inhibit learning or performance.

One of the first attempts to organize the varieties of learning was Melton's (1964) proposal of a simple taxonomy based primarily on clusters of tasks investigated by groups of researchers. The categories, roughly ordered by the complexity of the learning act, were *conditioning, rote learning, probability learning, skill learning, concept learning,* and *problem solving.* This general scheme has been given an updated treatment by Estes (1982) who examined conditions that facilitated and inhibited these and related classes of learning, and looked for evidence of individual differences in each class.

A task-based scheme was also the basis for learning taxonomies proposed by Jensen (1967) and Gagné (1965, 1985). Jensen proposed a three-faceted taxonomy (similar in some ways to Guilford's structure of intellect model): a *learning type* facet incorporated Melton's seven categories; a *procedures* facet indicated variables such as the pacing of the task, whether the task consisted of spaced or massed practice, stage of learning, and the like; and a *content* facet indicated whether the task consisted of verbal, numerical, or spatial stimuli. Jensen

proposed that his taxonomy could be used as an aid in interpreting some research findings, such as why arbitrarily selected learning tasks do not intercorrelate very highly (answer: because they do not share any facet values). He hoped that his taxonomy would suggest a more systematic approach to selecting learning tasks for future studies, but there is not much evidence that researchers have subsequently followed his suggestions.

Gagné's taxonomy (1965, 1985), on the other hand, has been widely taught and put to use in the area of instructional design (Gagné and Briggs, 1979). Gagné proposes five major categories of learned capabilities based on a rational analysis of common performances characteristics. *Intellectual skills* (procedural knowledge) reflect the ability to use rules; this capability in turn depends on the ability to make discriminations and to use concepts, and rules themselves combine to form higher-order rules and procedures. *Cognitive strategies* (executive control processes) reflect the ability to govern one's own learning and performance processes. *Verbal information* reflects the ability to recall and use labels, facts, and whole bodies of knowledge. *Motor skills* and *attitudes* are two additional learned capabilities Gagné includes to round out the list.

These categories serve various purposes. During task analysis, they assist the investigator in defining and analyzing instructional objectives and in evaluating an instructional system to determine whether its objectives have been met. For example, if the goal is to have the student acquire a conceptual skill, then the objective that the student be able to *discriminate* one thing from another may be indicated. In the design phase, the categories suggest different approaches for delivering instruction, since, according to Gagné, the five capabilities differ in the conditions most favorable for their learning. For example, with verbal information, order is not important but providing a meaningful context is, whereas for motor skills, providing intensive practice on part skills is critical.

All these taxonomic systems, Gagné's in particular, are beneficial, but it is important to acknowledge their limitations. One problem inherent in this approach is the degree to which it is subject to imprecision, which makes for communication difficulties and violates one of the main motivations for developing the taxonomy in the first place. Without a strong model of learning requirements in a task, and without a foundation of empirical relationships, task analysis is still primarily an art rather than a technology.

A second major problem with the rational approach was apparent to Melton (1964, 1967), who, in fact, argued that it be abandoned. The problem is that a taxonomic scheme based primarily on a rational analysis of task characteristics will only incidentally include actual psychological process dimensions. And presumably the process dimensions are what govern the most important aspect of the taxonomy: information regarding predicted task-to-task generality. Melton suggested that while the task-based approach might be initially useful, it was preferable ultimately to base the taxonomy on process characteristics rather than "a mish-mash of procedural and topographic (i.e., perceptual, motor, verbal, 'central') criteria" (p. 336). Although it was preliminary at that time to have actually suggested replacements to the task-based categories, we will show later how cognitive science now provides suggestions for what they might be.[1]

Correlational Taxonomies

A second approach, one less commonly used in the domain of learning skills, has been primarily empirical. The history of individual-differences research can be seen largely as an attempt to develop taxonomies of intelligence tests based on performance correlations (Thurstone, 1938), and there have been some attempts to develop similar taxonomies of learning tasks (Allison, 1960; Malmi, Underwood, and Carroll, 1979; Stake, 1961; Underwood, Boruch, and Malmi, 1978).

The correlational approach has one critical advantage over the rational approach as a means for taxonomy development: it directly addresses the issue of the transferability of skills among tasks. That is, if we know that performance on learning task X is highly correlated with performance on task Y, then a natural proposal is that a high proportion of the skills required by task X are also required by task Y. Further, training on task X should transfer at least somewhat to task Y. Thus patterns of correlations among performances on learning tasks could in principle be the basis for the construction of a taxonomy of learning skills.

A very closely related idea—that individual differences investigations could serve as acid tests in constructing general theories of learning—was developed by Underwood (1975). His proposal was that if a theory assumed some mechanism, and the mechanism could be measured in a context outside that in which it was initially developed, then the viability of the mechanism could be tested by correlational analysis.

These ideas were applied in an ambitious investigation that examined the intercorrelations among a wide variety of verbal memory tests (Underwood, Boruch, and Malmi, 1978). The purpose was to determine whether theoretical notions developed in the general (nomothetic) learning literature, such as the idea that memories have imaginal and acoustic attributes, or that recognition processes are distinct from recall processes, could be verified with an individual-differences analysis.

Memory-task stimuli were primarily words. In some tasks words were randomly selected, but in others words were chosen to elicit particular psychological processes. For example, concrete and abstract words were mixed, under the assumption that recall differences would reflect the degree of imagery involvement. Words were embedded in various kinds of memory tasks (paired associates, free recall, serial recall, memory span, frequency judgment). It was expected that clear word-attribute factors would emerge, thus supporting certain theoretical notions regarding properties of memory; however Underwood and colleagues discovered two somewhat unanticipated results. First, most of the variance was due to general individual differences in associative learning; only a small percentage was due to any subject-task interaction. Second, the two factors that did emerge were not associated with word attributes, as might have been expected, but with type of task (free recall versus paired associates and serial learning); but even this apparently is not a robust task division. A follow-up study (Malmi, Underwood, and Carroll, 1979) found the same evidence for a general associative-learning factor, but the two extracted factors split tasks in a slightly different way (free recall and serial learning versus paired associates).

What is the implication for a taxonomy of learning skills? Association formation rate apparently is a general and perhaps fundamental learning parameter. It may be that further subtle distinctions could be made among types of association formation, but the evidence in both these studies suggests little practical payoff in searching for such distinctions.

Underwood and colleagues were primarily interested in memory per se, and thus their tasks represent a fairly narrow range of learning. A useful complement to their analysis would be a study that more systematically sampled learning tasks from something like Melton's or Gagné's taxonomy. In this regard we consider a pair of studies by Stake (1961) and Allison (1960), who administered a diverse variety of learning tasks to large samples of seventh graders and Navy recruits, respectively. Allison's learning tasks were four paired-associates tasks (verbal, spatial, auditory, and haptic stimuli); four concept-formation tasks (spatial and verbal stimuli); two mechanical assembly tasks consisting of a short study film followed by an assembly test; a maze-tracing task; a standard rotary-pursuit task; and a task that involved learning how to plot quickly on a polar coordinates grid. Stake's learning tasks were listening comprehension (repeated study-test trials of the same story), free recall (words, numbers), paired associates (words, dot patterns, shapes, numbers), verbal concept formation, and maze learning. In both studies a variety of aptitude tests were also administered.

The original analyses of these data were somewhat problematic (see Cronbach and Snow, 1977), but a reanalysis conducted by Snow, Kyllonen, and Marshalek (1984) using multidimensional scaling (MDS) revealed a number of dimensions by which the learning tasks could be organized. First, in both studies, *learning tasks varied systematically in complexity.* This was indicated by two findings: the learning tasks varied substantially (a) in the degree to which

[1] It is historically interesting that it was at Melton's 1963 conference that Fitts (1963) proposed a highly process-oriented taxonomy of psychomotor skills which was only much later adapted by Anderson (1983) as the basis for a cognitive learning theory.

performance on them correlated with measures of general intellectual ability, and (b) in how close to the center of the multidimensional scaling configuration they appeared. Centrality reflects the average correlation of a test with other tests in the battery and may be taken as a measure of complexity (Marshalek, Lohman, and Snow, 1983; Tversky and Hutchins, 1986). Snow and colleagues suggested that the complexity relationship could be due either to some tasks subsuming others in terms of process requirements or to increased involvement of executive control processes such as goal monitoring.

Second, in both analyses there was evidence for a *novel* versus *familiar* learning task dimension, which Snow and associates interpreted as supporting the classical distinction between fluid and crystallized intelligence (Cattell, 1971), but which might also be seen as supporting a distinction between inductive and rote learning. In the Allison analysis, the paired-associates tasks and some of the concept-formation tasks appeared on one side of the scaling configuration. The concept formation tasks so positioned were those which repeatedly used the same stimuli, thus enabling the successful use of a purely rote strategy. On the other hand, the assembly tasks and the novel plotting task, which required subjects to assemble a new solution procedure essentially from scratch, appeared on the opposite side of the configuration.

The MDS analysis of the Stake (1961) data (learning rate scores) similarly suggested a fluid-inductive versus crystallized-rote dimension. Listening comprehension, verbal paired-associates, and verbal free-recall tasks appeared on the crystallized side of the configuration. The verbal concept-formation task along with the spatial and number pattern paired-associates tasks, which were partially amenable to an inductive learning strategy (response patterns could, but did not have to be induced), fell on the fluid-inductive learning end.

The reanalysis by Snow and colleagues thus provides a number of ideas that could facilitate taxonomy development. In particular it suggests task complexity and learning environment (inductive-novel task versus rote-familiar task) dimensions. Does this suggest we ought to continue along these lines to develop a full taxonomy? Unfortunately, we see two problems with the approach. One is simply practicality. Because of the time and expense involved in collecting data on performance of learning tasks, which typically require many more subject hours than do other cognitive measures, there have not been the same kind of large scale empirical analyses of learning-task batteries as there have been of intelligence-test batteries (although data sets reviewed in Glaser, 1967, and Cronbach and Snow, 1977, could be reanalyzed along the lines of the Snow et al. approach). Even with the well-designed studies Snow and colleagues reanalyzed, there is considerable underdetermination of process dimensions because not enough varieties of learning tasks were administered by Stake and Allison. Thus, although the dimensions that are revealed in the reanalysis by Snow and colleagues are suggestive, they certainly do not seem a sufficient basis for proposing a taxonomy of learning skills. It might take more like a few hundred diverse learning tasks to be able to see something that might serve as the basis for a true full-blown taxonomy. Obviously, such a study would be prohibitively expensive.

A second problem with the correlational approach to taxonomy building is one inherent in a purely bottom-up approach to theory development. That is, on what basis should learning tasks be selected for inclusion in a battery that is to be analyzed? Factor-correlational structures or categories directly reflect the nature of the tasks included in the analysis and only those tasks, and thus the empirical approach is inherently analytic and in some sense conservative. Correlational analyses certainly may be useful for initial forays or purely exploratory work in suggesting underlying relationships among tasks that might not have been anticipated at the outset. But it cannot be complete in any sense. One cannot simply be careful to "sample a broad range of tasks." A sampling scheme for choosing tasks already implies a taxonomy. Clearly, some means for generating original taxonomic categories is required.

Information-processing Model-based Taxonomies

The two classes of learning taxonomies thus far discussed have their roots in two different schools of thought—behaviorism in the case of rational taxonomies, psychometrics in the case of the empirical-correlational taxonomies—that historically precede modern cognitive psychology. One unfortunate side effect of the cognitive revolution had been a decline of interest in learning phenomena. Until the mid 1960s, when behaviorism was still largely predominant, learning issues held center stage. With the subsequent rise of cognitive psychology and the information-processing perspective, theories of memory and performance came to dominate. Only recently has there been a rather sudden and dramatic upsurge of interest in learning from an information-processing perspective. Although many of the same issues remain, these second looks at learning through newer theories (Anderson, 1983; Rosenbloom and Newell, 1986; Rumelhart and Norman, 1981) have resulted in a richer theoretical picture of learning phenomena.

Corresponding to this rise of interest in learning, there have been proposals for model-based categories or taxonomies of learning types. These attempts differ from the correlational taxonomies in that they have not yet been completely validated, at least not as taxonomies of learning skills. However, we do see correspondences between some of the dimensions that have emerged in the correlational analyses and some of the proposed learning mechanisms and categories, which we will point out as we go along. The model-based taxonomies differ also from the rational taxonomies in that they arise not simply from speculation and rational task analysis (although they certainly incorporate such methods) but from systematic information-processing models of learning that have been demonstrated to be specified to a degree of precision sufficient for implementation as running computer programs. Thus taxonomies in this category are those investigations that have entailed the use of computer simulation of learning processes as a means of developing learning theory.

One model-based taxonomy is suggested by Anderson's (1983) ACT* (Adaptive Control of Thought) theory. The theory proposes two fundamental forms of knowledge. *Procedural knowledge* (knowledge how) is represented in the form of a production system, a set of if-then rules presumed to control the flow of thought. *Declarative knowledge* (knowledge that) is represented in the form of a node-link network of propositions, which are presumed to embody the content of thought.

The ACT* theory, in its most recent formulation (Anderson, 1983; 1987a), specifies three basic types of learning, one to accommodate declarative (fact) learning, one specific to procedural learning, and one applicable to both types. Learning in declarative memory is accomplished solely by the probabilistic *transfer* to long-term memory of any new proposition (that is, a set of related nodes and links) that happens to be active in working memory. It is worth noting that the finding of Underwood and colleagues (1978) of a broad and general associative-learning factor lends empirical support to Anderson's claim for a single declarative-learning mechanism.

A second learning mechanism, *knowledge compilation,* accounts for procedural learning. Knowledge compilation actually consists of two related processes. Learning by *composition* is the collapsing of sequentially applied productions into one larger production. This corresponds to the transition from step-by-step execution of some skill to one-pass (all-at-once) execution. Learning by *proceduralization* is a related process in which a production becomes specialized for use in a particular task. This corresponds to the transition from the use of general problem-solving skills to tackle novel problems to the employment of task-specific skills, tuned to the particular problem at hand. Anderson's third learning mechanism, *strengthening,* operates somewhat analogously to the traditional learning principle of reinforcement. Both facts and procedures are presumed to get stronger and hence more easily and more reliably retrieved, as a function of repeated practice.

To appreciate Anderson's theory, it is important to note that it models the dynamics of skill transition, and is not simply a list of the different ways in which learning can occur or a categorization of learning tasks. The basic idea is that upon initial exposure to novel material, such as a geometry or computer-programming lesson, the learner first engages in declarative learning, forming traces of the various ideas presented. Then, when given problems to solve later in the lesson, the learner employs very general methods such as analogy, random search, or means-ends analysis, which operate on the declarative traces

to achieve solution. Employing these very general methods is cognitively taxing in that it severely strains working memory (to keep track of goals and the relevant traces), and thus initial problem solving is slow and halting. But portions of the process of using these general methods and achieving particular outcomes (some of which actually lead closer to solution) are automatically *compiled* while they are being executed. This is the procedural learning component. The learner essentially remembers the sequence of steps associated with solving a particular problem, or at least parts of the problem. Then when confronted with the problem again at some point in the future, the learner can simply recall that sequence from memory, rather than have to rethink the steps from scratch. With practice on similar problems, the compiled procedure is *strengthened,* which produces more reliable and faster problem solving. With continued practice, the skill ultimately is automatized in that it becomes possible to execute the skill without conscious awareness and without drawing on working memory resources.

Again there may be a correspondence between an individual-difference dimension and a distinction implicit in the model-based taxonomy. Snow and colleagues' novel learning tasks, presumed to tap fluid intelligence, may be likened to the novel learning situations that Anderson studies, which presumably tap very general problem-solving skills. On the other side, Snow and colleagues' familiar learning tasks, which call on crystallized skills, can be characterized in ACT* terms as engaging the declarative learning mechanism or involving the retrieval of already compiled procedures. It is noteworthy that despite rather major differences in methodology inherent in the individual-differences versus model-based approach, there is some convergence in the categories of learning skill. Although Anderson (1983, 1987a) views the emergence of the learning dimension as the result of the transition of skill, rather than perhaps as an array of fundamentally different kinds of learning tasks, there is a basic compatibility between the conclusions of the research approaches.

A second approach to building a model-based taxonomy is based on an integration of the literature from the artificial intelligence subspecialty of machine learning. Investigators have proposed taxonomies of research in machine learning (Carbonell, Michalski, and Mitchell, 1983; Michalski, 1986; Langley, 1986; Self, 1986), and there even exists something of a consensus in the field regarding the categories in the taxonomy.

One dimension of machine-learning research particularly relevant to our concerns here is *learning strategy,* which Michalski (1986) defines as the *type of inference* employed during learning, and which he characterizes as follows:

> In every learning situation, the learner transforms information provided by a teacher (or environment) into some new form in which it is stored for future use. The nature of this transformation determines the type of learning strategy

used. . . . These strategies are ordered by the increasing complexity of the transformation (inference) from the information initially provided to the knowledge ultimately required. Their order thus reflects increasing effort on the part of the student and correspondingly decreasing effort on the part of the teacher (p. 14).

It is interesting that the classification of machine-learning research yields such a nice process classification and thereby seems promising as a realization of Melton's ultimate hopes for a taxonomy of learning. The kinds of inferencing strategies Carbonell and colleagues and Michalski suggest are listed in Table 4-1. (We have added an additional category, "learning by drill and practice," to the list because we use the list as the basis for one of the proposed taxonomy categories, and it is convenient to denote that here.) Note that while there may be some similarity between the categories of Carbonell et al. and Michalski and those proposed by Melton, Gagné, and others, the basic difference is that in the Carbonell-Michalski system the underlying motivation for distinctions is necessarily the existence of differences in cognitive processing requirements. We will return to a more thorough discussion of these categories in the next section.

We believe that Anderson's and Carbonell-Michalski's model-based attempts to propose varieties of learning represent a considerable advance beyond either the rational or correlational taxonomies and go a long way in abating some of the most severe criticisms of earlier taxonomies. Yet all three approaches yield ideas on the varieties of learning skills that might be fruitfully synthesized. The remainder of the chapter will represent our initial attempt to integrate these ideas.

A PROPOSED TAXONOMY OF LEARNING

Thus far we have discussed why a taxonomy of learning is important, and what others have done in the way of proposing taxonomies. We now proceed to propose a taxonomy based on a synthesis of some of the ideas just reviewed, with an eye toward two major objectives. First, the taxonomy should be useful as a learning task–analysis system. That is, it should be useful in answering questions like what are the component skills involved in learning to disassemble a jet engine, operate a camera, program a computer, or make economic forecasts? Second, the taxonomy should serve to focus our research. Specifying the ways people learn may suggest where we ought to be expending more research energy. We do not see this as dictating research directions, as some critics of psychological taxonomies have suggested (Martin, 1986), but as suggesting potentially high-payoff research directions. For example, we already know much about declarative learning, such as what kinds of individual differences to expect and the relation of declarative learning to other cognitive skills. We know considerably less about procedural learning skills. The taxonomy may

Table 4-1 Learning Strategies from a Taxonomy of Machine-Learning Research

Rote learning. Learning by direct memorization of facts without generalization.

Learning from instruction (advice taking; learning by being told). The process of transforming and integrating instructions from an external source (such as a teacher) into an internally usable form.

Learning by deduction

- Knowledge compilation. Translating knowledge from a declarative form that cannot be used directly into an effective procedural form; for example, converting the advice "Don't get wet" into specific instructions that recommend *how* to avoid getting wet in a given situation.
- Caching. Storing the answer to frequently occurring questions (problems) in order to avoid a replication of past efforts.
- Chunking. Grouping lower-level descriptions (patterns, operators, goals) into higher-level descriptions.
- Creating macro-operators (*composition*). An operator composed of a sequence of more primitive operators. Appropriate macro-operators can simplify problem solving by allowing a more "course-grained" problem-solving search.

Learning by drill and practice. Refining or tuning knowledge (or skill) by repeatedly using it in various contexts and allowing it to strengthen and become more reliable through generalization and specialization.

Inductive learning. Learning by drawing inductive inferences (a mode of reasoning that starts out with some assertions, e.g., specific observations, and concludes with more general and plausible assertions, i.e., hypotheses explaining the initial assertions) from facts and observations obtained from a teacher or an environment.

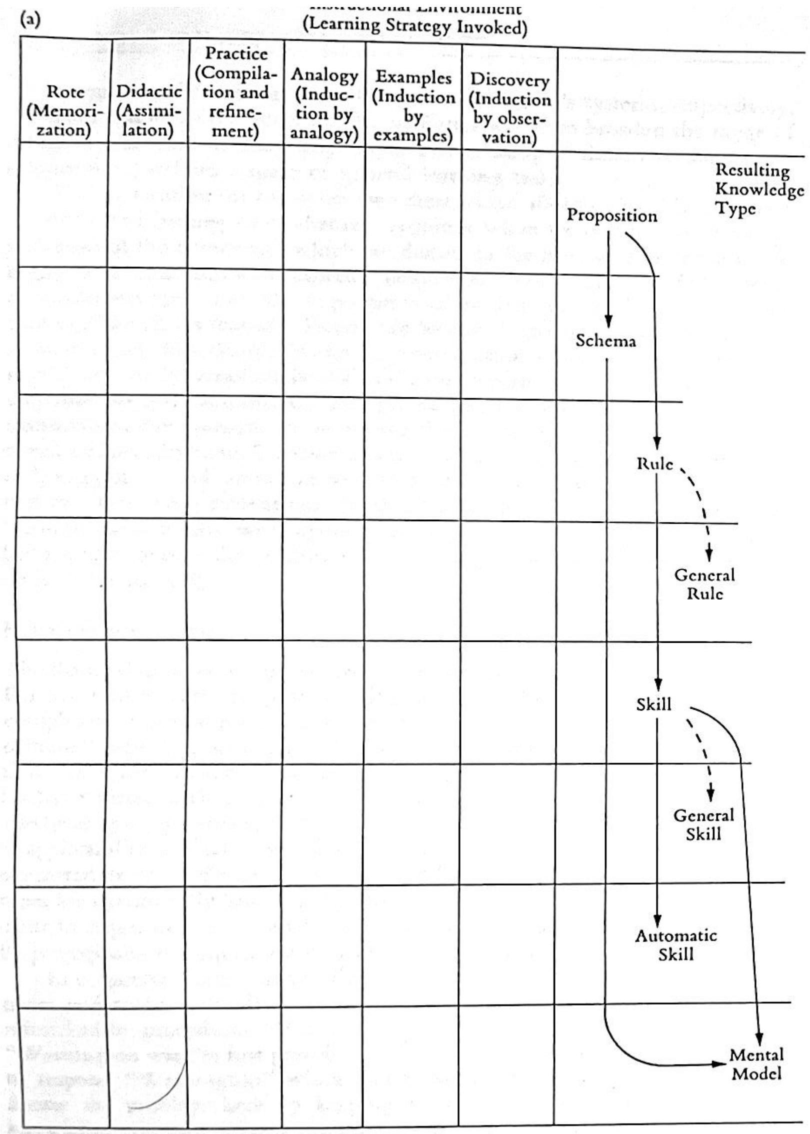
- Learning by analogy. Mapping information from a known object or process to less known but similar one.
- Learning from examples. Inferring a general concept description from examples and (optionally) counterexamples of that concept.
- Learning from observation and discovery (learning without a teacher; unsupervised learning). Constructing descriptions, hypotheses, or theories about a given collection of facts or observations. In this form of learning there is no a priori classification of observations into sets exemplifying desired concepts.

Note: All categories except *"learning by deduction"* are from Carbonell et al. (1983); *"learning by deduction"* is from Michalski (1986). The definitions are taken from the glossary in Michalski et al. (1986). Note that *"learning by drill and practice"* was not a category included by Carbonell et al. (1983) or Michalski (1986), but we included it in the taxonomy, and thus for economy we describe it here.

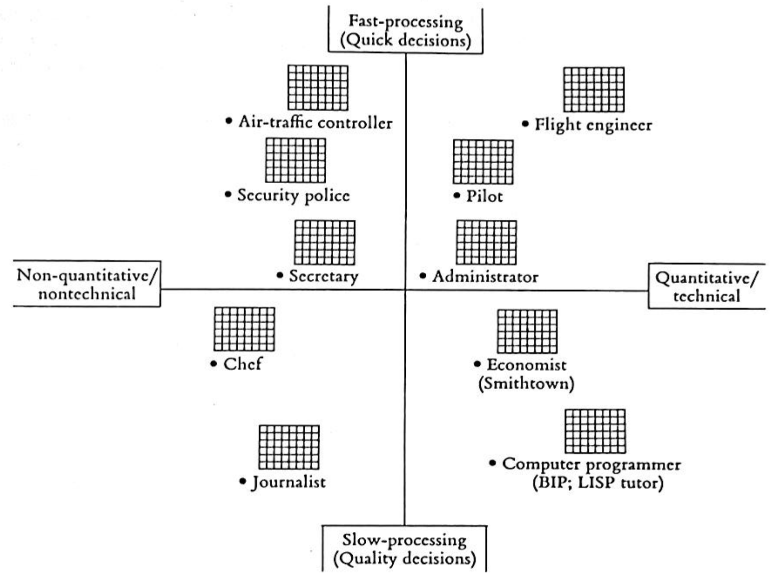
pinpoint other learning skills on which research attention may productively be focused.

We have selected four dimensions, illustrated in Figure 4-1, as particularly important in classifying learning skills. The two dimensions shown in Figure 4-1a—*knowledge type* and *instructional environment*—are motivated primarily by

(a)



(b)



(c)

Holistic processing ⟷ Serial processing
Active/impulsive orientation ⟷ Passive/reflective orientation
Systematic/punctilious approach ⟷ Haphazard/exploratory approach
Theory-driven (top-down) ⟷ Data-driven (bottom-up)
Spatial representation ⟷ Verbal representation
Deep processing ⟷ Superficial processing
Low internal motivation ⟷ High internal motivation

Figure 4-1 Learning skills taxonomy. a. Environment-by-knowledge type matrix: cell entries would be various learning tasks. b. Environment-by-knowledge type matrices plotted in a hypothetical two-dimensional domain-space: proximal matrices should show relatively greater transfer among parallel learning skills. c. Suggested learning styles that might interact with other taxonomy dimensions in determining what learning skill a particular learning task measures.

our discussion of Anderson's and Carbonell-Michalski's systems, respectively, although Gagné's ideas on learned capabilities served to broaden the range of categories included in knowledge type. The crossing of these two dimensions (Figure 4-1a) defines a space of general learning tasks.

The motivation for the other two dimensions, illustrated in Figure 4-1b,c —*domain* and *learning style*—became apparent when we began examining applications of the taxonomy, which we discuss in the next section of the paper. Figure 4-1b illustrates a hypothetical *domain-space* as the crossing of the degree of quantitativeness and the importance of quality versus speed in decision making. The idea is that any domain can be located in such a space, and that the set of learning skills defined by the first two taxonomy dimensions (Figure 4-1a) may prove to be empirically distinct from parallel learning skills in other domains. We represent this idea in Figure 4-1b by scattering knowledge type by instructional environment matrices over the domain space, for various occupational-training domains. The two dimensions portrayed in the domain space are only suggestive, and are meant only to express how domain interacts with the first two taxonomy dimensions. Finally, Figure 4-1c lists a variety of possible learning styles, which, we propose, must be considered in conjunction with the first three taxonomy dimensions in determining what skills are being tapped by a particular learning task.

Knowledge Type

The distinction between declarative and procedural knowledge is fundamental. Further refinements are possible: declarative knowledge can be arrayed by complexity, from propositional knowledge to schemas (packets of related propositions). Similarly, procedural knowledge can be arrayed from simple productions, to skills (packets of productions that go together), to automatic skills (skills executed with minimal cognitive attention). Productions and skills can also be arrayed by generality, from a narrow (specific) to a broad (general) range of applicability. A final knowledge type is the mental model, which requires the concerted exercise of multiple skills applied to elaborate schemata. Knowledge types are dynamically linked: acquisition of a set of propositions may be prerequisite to acquisition of a related schema or to a procedural skill; both in turn may be prerequisite to acquisition of some mental model.

In cognitive science circles, the declarative-procedural distinction is sometimes said to be formally problematic in that declarative knowledge can be mimicked by procedures (Winograd, 1975). One can declaratively know that "Washington was the first president", alternatively, one can have the procedure to respond "Washington" when asked "who was the first president?" We finesse the problem here by keeping close to an operational definition of knowledge type: we define knowledge in terms of how it is tested. Declarative

knowledge can be probed with a fact recognition test (sentence recognition, word matching), or, in the case of schemata, with clustering and sorting tasks (Chi, Feltovich, and Glaser, 1981). Procedural knowledge requires a demonstration of the ability to apply the knowledge to predict the output of some operator (operator tracing) or to generate a set of operators to yield some output (operator selection). Possession of skills and automatic procedures may be operationally determined by examining the degree of performance decrement under imposition of secondary tasks (Wickens, Sandry, and Vidulich, 1983) or through other methods of increasing processing demands (Spelke, Hirst, and Neisser, 1976; Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). Possession of an appropriate mental model might require testing performance on a complex simulation of some target task. Table 4-2 gives an illustrative (not exhaustive) list of tests for the various knowledge types.

Instructional Environment

Instruction delivered in a classroom setting or even on a computer will inevitably provide the student with opportunities to incorporate the material in multiple ways. Real instruction occurs in a diverse environment from the standpoint of student control versus teacher control and consequently in the kinds of inferences students are required to make. Nevertheless, it is useful to differentiate instructional environments in a local sense: it should be possible to tag a specific instruction segment as to the form in which it is delivered and the kinds of inference processes or learning strategies it is likely to invoke. Following Carbonell et al. and Michalski (Table 4-1) we propose to characterize local instructional environments according to the amount of student control in the learning process. At one end, rote learning (such as, memorizing the multiplication table) involves full teacher control, little student control. Didactic learning (by textbook or lecture), learning by doing through practice and knowledge compilation, learning by analogy, learning from examples, and learning by observation and discovery offer successively more student control and less teacher control.

Note that we modify the Carbonell-Michalski list slightly by combining their learning-by-deduction (compilation) category with a learning-by-refinement category (suggested to us by W. Regian, personal communication, May 4, 1987). What we are pinpointing is the ability to refine one's skill (by strengthening, generalization, and discrimination) based on feedback following performance. Before one is engaged in this kind of learning, we assume the skill has already been acquired (perhaps in a rote fashion) and compiled, and is now at the phase of being refined. But because compilation and refinement are probably hopelessly intertwined in actual learning contexts, we combine them into a single learning-by-doing (practice environment) category.

Table 4–2 Sample Tests for the Various Knowledge Types

Knowledge type	Type of test	Sample item
Proposition	*Sentence verification*	*"AND yields High if all inputs are high, Low otherwise—True or False?"*
	Stimulus matching	*"AND D—Match or Mismatch?"*
	Paired associates	*"Which symbol is associated with AND?"*
	Free recall (components)	"What are the different types of logic gates?"
Schema	*Free recall (structure)*	"Reproduce the circuits you just studied"
	Sorting	"Sort the circuits into categories"
	Classification	"Pair circuit diagrams with these devices"
	Sentence completion/cloze	"AND yields ___ if all ___ are ___"
	Lexical decision	"XAND is a legal logic gate—True or False?"
Rule	*Operator tracing*	Determine output of logic gate (AND, HIGH, LOW) = ?
	Operator selection	Choose an operator to achieve a result (?, HIGH, LOW) = HIGH
General rule	*Transfer of training*	Learn and be tested on other kinds of logical relations such as those introduced in symbolic logic
Skill	*Multiple operator tracing/selection*	Trace through (or select) a series of linked logic gates in a circuit [could also use hierarchical menus methodology]
General skill	*Transfer of training*	Learn and be tested on constructing or verifying logical proofs
Automatic skill	*Dual task*	Trace logic gates while monitoring a secondary signal
	Complexity increase	Trace logic gates that become increasingly complex
Mental model	*Process outcome prediction*	Troubleshoot a simulated target task; "walk-through" performance test

Note: Sample items are tests that might be administered to a student finishing a lesson on logic gates as part of a course in electronics troubleshooting (see, for example, Gitomer, 1984).

Domain (Subject Matter)

The inclusion of subject matter as a taxonomy dimension reflects the fact that much of learning has a strong domain-specific character. One can be an expert learner in one domain and a poor learner in another. Certainly there is some generality in learning skills over domains. Glaser, Lesgold, and Lajoie (in press) suggest that metacognitive skills might be fairly general. But even here, there is little evidence that metacognitive skill in mathematics (Schoenfeld, 1985) predicts metacognitive skill in writing (Hayes and Flower, 1980).

It is appropriate to ask the question of the topic range over which some general learning skill is likely to be useful. It may be that the degree to which a subject matter taps quantitative or technical knowledge, and the degree to which it taps verbal knowledge captures some of the transfer relations among academic subjects. The degree of social involvement may also play a role, especially when considering the universe of occupational training courses rather than simply academic training. As is suggested in Figure 4-1b, it may be that the relative importance of speed versus quality in decision making is a critical domain dimension. But again, the dimensions portrayed in Figure 4-1b are only meant to be suggestive.

More generally, we envision a complete *domain-space*. The underlying dimensionality of such a space could be discovered through a study of the similarity (either judged or as shown in transfer of performance relations) among all jobs, courses, or learning experiences in any specifiable universe of interest, and could be represented as a multidimensional scaling of the jobs or courses so rated. An empirically determined domain-space would specify the likelihood that (or the degree to which) a particular taxonomic skill, defined by the environment and the knowledge type, would transfer to or be predictive of a parallel skill (that is, one defined by the same environment and knowledge type) in another domain. Proximal domains, in the multidimensional space, would yield high transfer among parallel skills; distal domains might yield only minimal transfer. For example, assuming the importance of the quantitative dimension, skill in learning mathematics propositions through didactic instruction might predict skill in learning physics propositions through instruction; but neither may be related to the ability to learn history propositions through instruction.

Learning Style

All sorts of subject characteristics — aptitudes, personality traits, background experiences — affect what is learned in an instructional setting. But we focus on characteristics of the learner's preferred mode of processing, or learning style, because our primary concern is characteristics over which the instructional designer may exercise control. Because style implies a choice by subjects as to how to orient themselves toward the learning experience, it should be manipulable through instruction.

A considerable literature on cognitive style exists (Messick, 1986). Among those that have received the most attention are field dependence-independence (Goodenough, 1976) and cognitive complexity (Linville, 1982), but these are now presumed to primarily reflect ability (Cronbach and Snow, 1977; Linn and Kyllonen, 1981). Impulsivity-reflectivity (Baron, Badgio, and Gaskins, 1986; Meichenbaum, 1977) more clearly fits our criteria for inclusion in the taxonomy, in that it is malleable: subjects can be trained to be more reflective in problem solving, and this improves performance. Other styles we consider in our analyses of learning environments are holistic versus serial processing, activity level, systematicity and exploratoriness, theory-driven versus data-driven approaches, spatial versus verbal representation of relations (Perrig and Kintsch, 1984), superficial versus deep processing, and low versus high internal motivation. Some dimensions may affect learning outcomes quantitatively: active students may learn more. Others may affect outcomes qualitatively: spatial versus verbal representations will result in different relationships learned.

Cognitive style may interact with other taxonomy dimensions in determining what learning skill is being tapped in instruction. A study by Pask and Scott (1972) which identified holist versus serialist processing styles illustrates this interaction. In this study, serialists, those who focus on low-order relations and remembering information in lists, were contrasted with holists, who focus on high-order relations and remembering the overall organization among items to be learned. Pask and Scott showed that presenting a learning task (that is, learning an artificial taxonomic structure) in a way that matched the learner's style resulted in better overall learning. A critical point for this discussion is that the presentation of material should tap different skills for subjects who differ on this style dimension. Presenting a long list of principles may be a difficult memory task for serialists, who attempt to memorize each relationship presented. For holists, the same task may tap conceptual reorganization skill rather than memorization skill.

Summary

The first three dimensions of the taxonomy define a space of learning tasks (Figure 4-1a set in the domain-space of Figure 4-1b). Each cell represents a task that teaches a particular subject matter (such as physics principles: Newton's second law), by a particular means (for example, by analogy), resulting in a particular kind of knowledge outcome (for example, a schema). A particular taxonomic learning skill then may be defined by performance on a particular taxonomic learning task. There will be interactions among dimensions: some subject matters lend themselves more readily to certain kinds of knowledge outcomes. For example, propositions are emphasized in nonquantitative fields; procedures are the focus in quantitative fields. And knowledge outcomes covary with instructional method: we more commonly learn propositions than procedures by rote.

As an illustration of some of these ideas, consider the instructional goal of teaching the concept of *electric field* (Glynn et al., in press). A *rote* approach might be to have students simply memorize the definition: "an electric field is a kind of aura that extends through space." A *didactic* approach might specify that students read the definition embedded in the context of a larger lesson, then to have the student demonstrate understanding by having him or her paraphrase the definition. The difference between the two approaches could be reflected in the way in which the knowledge was tested. The appropriate rote test would be verbatim recognition or recall; the appropriate instruction test would be paraphrased recognition or recall.

The electric field concept could be taught by having students *practice* using it: following a discussion of properties of force such as how an electrical force holds an electron in orbit around a proton, students would be given an opportunity to solve problems that made use of the concept. One could also lead students to *induce* the concept, by pointing out how it is *analogous* to a gravitational field, by providing them with *examples* and counterexamples, or by having them *discover* it with a simulator, or in a laboratory.

Unlike the first three dimensions, the fourth dimension — learning style — refers to characteristics of the person rather than the environment. Inclusion of the learning style dimension is an admission that providing a particular kind of environment guarantees neither the kind of learning experience that will result nor the kind of learning skill being tapped. Interactions exist between person characteristic and instructional treatment (Cronbach and Snow, 1977, especially Chapter 11), and thus, as we tried to illustrate in the example on holist versus serialist processing, the style engaged at the time of learning and testing will partly determine what learning skill is being measured.

APPLYING THE TAXONOMY: THREE CASE STUDIES

In this section of the chapter we want to consider how the learning taxonomy might facilitate the development of indicators of learning skill in actual practice. We consider this a kind of test run for the taxonomy. Having proposed a taxonomy, we will now demonstrate how it might be applied. We discuss three computerized instructional programs, each of which includes some capability for determining what and how students are learning. We suggest ways in which additional learning indicators might be generated in light of our taxonomy.

We see the taxonomy playing two roles here. One, although not the focus of the chapter, is to help us classify instructional programs. By our taxonomy, similar programs are ones that teach the same type of knowledge (propositions, skills, and so on), provide the same instructional environment (rote, discovery), teach the same domain material such as computer programming, economics), and encourage the same kind (style) of learner interaction (reflectivity, holistic processing, and so on). Programs are dissimilar to the degree that they mismatch on these dimensions. An important part of our discussion of the three tutoring systems then is to indicate at least informally what learning skills are being exercised, and to what degree.

The second and, for current purposes, more important role for the taxonomy is to assist us in thinking more broadly about learning skills and outcomes. The taxonomy with its specified methods and tests can pinpoint what potentially important learning events are simply not being measured by existing instructional programs. We can imagine generating alternative instructional programs by varying the degree to which different kinds of learning skills are exercised.

The three programs we discuss in this section are intelligent tutoring systems, and so we begin by providing a few preliminary remarks on their general organization.

General Comments on Intelligent Tutoring Systems

Figure 4-2 illustrates the components of a hypothetical and somewhat generic intelligent tutoring system. In this system, the student learns by solving problems, and a key system task is to generate or select problems that will serve as good learning experiences.

The system begins by considering what the student already knows—the *student model*—what the student needs to know—the *curriculum*—and what curriculum element (lesson or skill) ought to be instructed next—the *teaching strategy*. From these considerations the system selects (or generates) a problem, then either works out a solution to the problem (with its *domain expert*), or simply retrieves a prepared solution. The program then compares its solution to

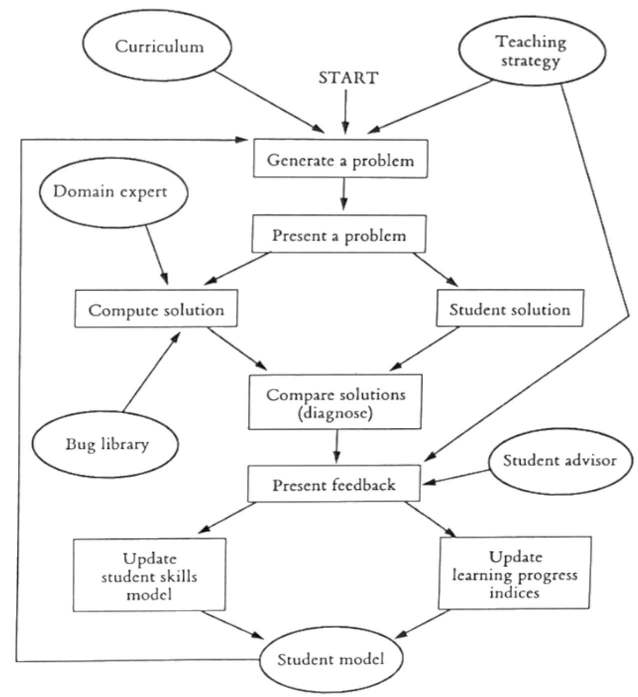


Figure 4-2 Components of a generic intelligent tutoring system. (Boxes represent decisions the program makes; ellipses represent knowledge bases the program consults.)

one the student has prepared and performs a diagnosis based on the differences between the solutions.

The program provides feedback based on student-advisor considerations such as how long it has been since feedback was last provided, whether the student was already given a particular bit of advice, and so forth. After this, the program updates both the student skills model (a record of what the student knows and does not know) and increments learning progress index counters. These updating activities modify the student model, and the entire cycle is repeated, starting with selecting or generating a new problem.

Not all ITSs include all these components, and the problem-test-feedback cycle does not adequately characterize all systems. But this system fairly describes many existing ITSs and perhaps most interactions with human tutors. Thus an examination of the components of the generic tutor should yield some ideas on how learning progress and the current status of the learner may be indicated. Note that much of this information is contained in the dynamic student model. We now discuss three instantiations of this generic tutor.

BIP: Tutoring Basic Programming

General System Description The Basic Instruction Program (BIP) was one of the first operational intelligent tutoring systems (Barr, Beard, and Atkinson, 1976; Wescourt et al. 1977).[2] BIP teaches students how to write programs in the language BASIC by having the student solve problems that get progressively more difficult. The system selects problems according to what the student already knows (based on past performance), which skills it believes ought to be taught next, and its understanding of the skills required by the problems in its problem bank.

The BIP architecture is consistent with the generic tutor. Its *Curriculum Information Network* represents all the skills to be taught and the relations among them. Skills are represented quite narrowly, for example, "initialize a counter variable" or "print a literal string." The relations specify whether skills are analogous to other skills, whether they are easier or harder or at the same difficulty level as other skills, and whether there are any prerequisite skills. As examples, (a) printing a numeric literal (or constant) is considered conceptually analogous to, (b) but also easier than, printing a string literal; (c) both are considered easier than printing a numeric variable; (d) printing a numeric literal is considered a prerequisite to printing the sum of two numbers.

A programming *task* is represented in terms of its component skill requirements. For example, a BIP task might ask the student to compute and print out the number of gifts sent on the twelfth day of Christmas, given that on the first day 1 gift was sent, on the second day 1 + 2 gifts were sent, on the third day, 1 + 2 + 3 were sent, and so on. The student is expected to write a program that computes the sum of 1 + 2 + . . . + 12. Based on a task analysis conducted by BIP's authors, BIP knows that the component skills required for solving this particular problem are *initialize numeric variable, use for-next loop with literal as final value,* and so forth. Each task is assumed to tap a number of skills.

The BIP student model is a list of the student's status with respect to each of ninety-three skills in the curriculum. There are five discrete status levels: *unseen* (not yet seen a problem that required the skill), *trouble* (seen but has not

[2] Barr et al. developed BIP-I; Wescourt et al. developed its successor BIP-II. The two systems are fairly similar, but we assume the newer system where there are discrepancies.

solved a problem that required the skill), *marginal* (learned to a marginal degree), *easy* (not yet seen but an easy skill to learn), and *learned* (to a sufficient degree). After each problem, skill status is updated as a result of the student's self-evaluation and through two domain-expert–like components to BIP: a BASIC interpreter that catches syntax errors, and a solution evaluator that determines whether the program is producing correct output. Finally, BIP also provides a number of aids to the student. The student may request help (suggestions as to how to solve the problem), a model solution (such as, a flowchart), or a series of partial hints.

BIP selects problems by first identifying skills the student is ready for (ones that do not have any unlearned prerequisites) but that need work, which means (in order of priority) (a) skills students have had trouble with (from tasks they have quit), (b) skills analogous to learned skills, or (c) skills postrequisite to learned skills. It calls skills so identified *needed* skills. BIP then identifies a *task* with needed skills but no unlearned prerequisites.

If the student successfully solves the selected task, BIP updates the student model by crediting the associated task skills. If the student fails the problem or gives up (that is, requests a new task), BIP determines which skills to blame according to criteria such as the student's self-evaluation, whether the student already learned some of the skills or analogous ones, and whether any task skills or analogous ones are in an unlearned state.

There are a number of ways in which aptitude information guides problem selection. For the fast learner, if two skills are linked by difficulty (one is harder than the other), the system assumes that the easier one is not a needed skill; BIP also will select tasks with multiple needed skills. If the student is consistently having trouble, BIP opts for a slow moving approach and minimizes the number of needed skills introduced in a single task.

Learning Indicators Snow, Wescourt, and Collins (1986) collected aptitude and other personal data from twenty-nine subjects who had used BIP and performed a number of analyses on the relationships between those data and BIP variables. Table 4-3 shows the list of learning indicators used by Snow et al. We have divided the list into three categories: learning summary indices, learning activity variables, and time allocation variables.

The sample was too small to draw definitive conclusions about relationships, but there were some suggestive findings worthy of further pursuit. First, the *best* learning progress index seemed to be the slope of the number of skills acquired over the number of skills possible—that is, skills slope. Determination of best is based on two considerations: skills slope was most representative of other learning progress indices in that it had higher average intercorrelations with those indices (centrality), and it had higher average correlations with the learning activity variables (a validity of sorts). Particularly intriguing was that skills slope, along with a global achievement posttest, was more highly related

Table 4–3 Learning Indicators from BIP, the Programming Tutor

Learning summary indices
1. Number of problems seen
2. Mean time per problem
3. Number of skills acquired
4. Skills acquired per problem (slope, intercept, standard error)
5. Skills acquired per time on task (slope, intercept, standard error)
6. Skills acquired per skills possible (slope, intercept, standard error)

Learning activity variables
(counts of activities, to be divided by number of problems seen)
1. Student produces correct solution
2. Student has difficulty on the task (according to BIP)
3. Student admits not understanding the task
4. Student disagrees with solution evaluator
5. Student requests solution model
6. Student requests solution flow chart
7. Student requests model program
8. Student starts problem over
9. Student requests at least 1 hint before starting
10. Student requests at least 1 but not all hints
11. Student requests all hints (0–5 on a problem)
12. Student quits the problem
13. Student quits the problem after seeing all the hints
14. Student quits the problem without seeing any hints
15. Student tests different input cases after successful solution
16. Student tests different input cases after failed solution
17. Student uses BIP input data after failed solution
18. Student runs program parts rather than complete program
19. Student requests aid (model, help, hint) after an error

Time allocation variables*
21. Planning: Proportion of time spent before coding
22. Implementing: Proportion of time spent writing code
23. Debugging: Proportion of time spent debugging code

*Time on the tutor must fall into one and only one of the three time allocation portions.

to the activity variables than was the raw number of skills acquired. Snow and colleagues suggested this may have been due to skills slope capturing more about the progress of learning over time.

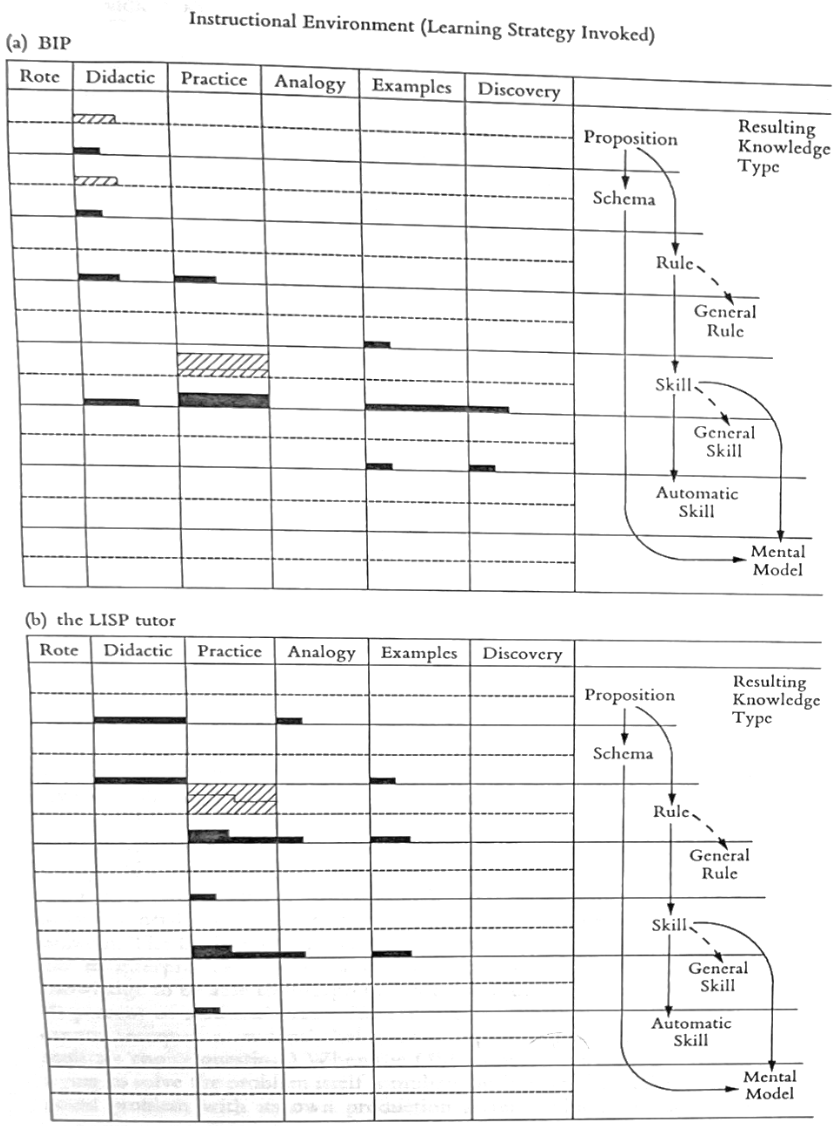
The second major finding concerned the role of the activity variables in predicting learning outcome. As it turned out, most of the tool-use indicators, such as requests for demonstrations, hints, and model solutions, were associated with poor posttest performance. Poor performers also spent more time debugging and less time planning than did others, and were more likely to quit the task or start over. In contrast, good performers requested fewer hints, spent more time implementing rather than debugging, and were more likely to test different cases after a successful run of their program (indicator 15). This may have reflected good students' desire to perform additional tests of their knowledge, perhaps to probe the boundaries of their understanding, even after passing the test.

Applying the Taxonomy In evaluating the BIP tutor with respect to the taxonomy, we ask two questions: (a) what learning skills does BIP exercise (that is, how can BIP be classified), and (b) how comprehensive are the indicators used by Wescourt et al. and Snow et al. in measuring students' learning skills and their learning progress?

To address the first question, consider a distinction between what is tested for and what is taught. BIP primarily tests for fairly specific skills in that virtually all its tests are of the multiple operator selection variety, meaning that students write programs. The posttest also undoubtedly taps some propositional schematic knowledge, but not extensively. Other knowledge outcomes could be tested for but they are not. BIP teaches skills by having students first read a text (*learning from instruction*, in taxonomy terminology), then apply the studied skills in a problem-solving context (*learning through compilation and learning by drill and practice*). Some students also request help and thereby engage in *learning from examples*. The good students also tend to invoke *observational learning* when they perform additional tests of their programs.

Figure 4-3a summarizes our assessment of (a) what skills are being exercised by BIP, indicated as the solid bar, and (b) what skills are being tested for, indicated as the striped bar. Bar size represents the proportion of time spent either engaging the learning skill (solid) or having the ability tested (striped), relative to engaging or testing other skills. It is important to keep in mind that this analysis is rather informal. We made some rough computations of the times engaged in the various activities, based on a review of the data on the learning indicators of Snow et al., and on the report of Wescourt et al. of some other summary statistics. Our analysis is meant to be merely suggestive. A more rigorous, systematic analysis of BIP could produce a precise breakdown, separately for each student, of the time spent exercising and testing various learning skills. Also note that Figure 4-3 indicates only the knowledge type and instructional environment dimensions. Domain is indicated in Figure 4-1b (computer programming is highly quantitative and technical, and the quality of decisions is emphasized). Learning style is not directly assessed in BIP.

An approach to the second question, concerning indicator comprehensiveness, is suggested by Figure 4-3a: which skills are being exercised and not tested? First, we can see that although students are learning rules, they are not tested for them. This could be remedied by including operator tracing or



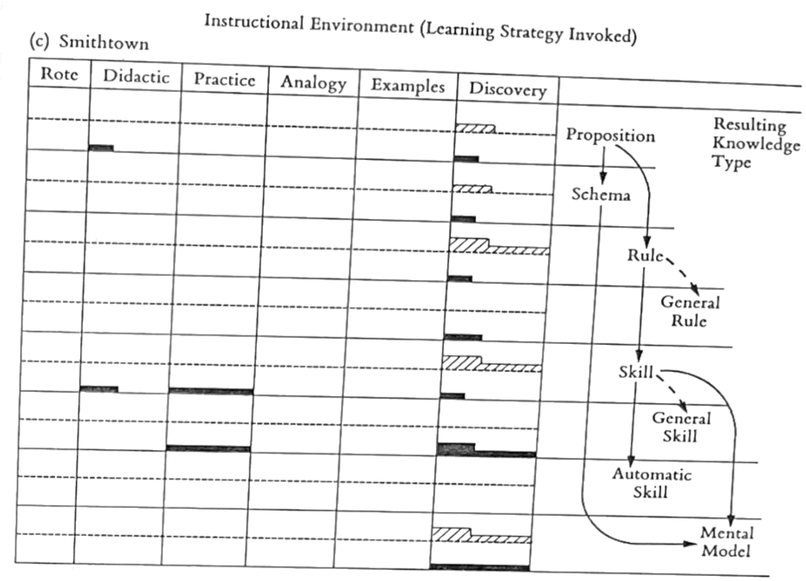


Figure 4-3 Learning activities profiles for (a) BIP, (b) the LISP tutor, and (c) Smithtown; solid bars represent the proportion of time the particular skill (defined by the environment-by-knowledge type cell task) is *exercised* by the tutor, relative to other skills; striped bars represent the proportion of time the skill is *tested*, relative to other skills.

selection tests. Second, students also are probably acquiring some general rules and skills regarding program-writing strategies, but BIP does not directly test for these. Transfer of training tests inserted into the program (as part of the curriculum) would help determine the generality of the skills learned in BIP. Third, students read text, and get tested on their knowledge during the posttest, but it would be possible to more directly test the propositional and schematic knowledge resulting from reading the text by administering sentence verification tests, sorting tasks, and the like (see Table 4-2). Finally, the task of writing programs is an operator selection task and thus is more difficult than a task that would require students merely to understand the workings of a program (an operator tracing task). Students may understand a program they are unable to write. The inclusion of a program understanding task would tap knowledge that would be missed otherwise and thus should enhance the accuracy of the student model.

In sum, BIP generates many indicators of student status and learning progress. Application of the taxonomy suggests a number of additional ways in which students' knowledge and learning skill could be assessed. Expanding the breadth of learning-skill probes should affect the overall quality of any intelligent tutoring system both in its role as a training device and as a research tool. The performance of an ITS with a student-modeling component is highly dependent on the quality of the student model insofar as the system's main job is to select problems that are of an appropriate level. Thus an ITS should improve with a better student model, and we made suggestions here for refining a student model. As a research tool, an ITS can serve as an environment in which to examine the interrelationships among learning skills and learning activities. Snow et al.'s analysis of BIP relied on a rich set of learning indicators. However we think that the taxonomy can be used to provide an additional psychological basis for expressing those indicators.

Anderson's LISP Tutor

General System Description Anderson and his research group have developed intelligent tutoring systems for geometry, algebra, and the programming language LISP. We focus here on the LISP tutor. Descriptions of the tutor are available (Anderson, Boyle, and Reiser, 1985), and thus we only summarize some of the main features of the system especially as they contrast with BIP.

The LISP tutor follows the generic architecture fairly closely. Students read some material in a textbook, but then go on to spend most of their time interacting with the program. The program selects problems, gives the student help or advice when asked, and interrupts if the student is floundering.

An innovation of the LISP tutor is its use of what Reiser, Anderson, and Farrell (1985) call the *model-tracing* methodology, the process by which the tutor understands what the student is trying to do while the student attempts to solve a problem. Whenever the student types in an expression (as part of a solution attempt) the tutor evaluates the expression as to whether it is the same as what the ideal student would type in, or whether it indicates a misconception (or *bug*). If a misconception is indicated, the tutor intervenes with advice.

For a tutor to analyze the student's response so microscopically, it has to know essentially every correct step and every plausible wrong step in every problem. The LISP tutor does not incorporate enough domain knowledge to be able to interpret *every* action a student might take, but it does have enough knowledge to be able to interpret all correct solutions and approximately 45 to 80 percent of students' errors (Reiser et al., 1985). (In cases where the tutor cannot interpret a student's behavior it typically probes the student with a multiple choice question.) When the LISP tutor poses a problem, it goes about trying to solve the problem itself, simultaneously with the student. It solves the posed problem with its own production system, which consists of approximately 400 production rules for correctly writing programs (Anderson, 1987b). It also solves the problem in various plausible incorrect ways, through the action of about 600 incorrect ("buggy") production rules. Determining what the student is doing is a matter of comparing student input with its internal production system results.

Learning Indicators The LISP tutor keeps a record of the student's status with respect to each skill being taught, where skills are the 400 correct production rules. An indicator of how well the student knows a rule is incremented when the student uses the rule correctly, and decremented when the student makes an error. Remedial problems may be selected to give a student experience in using a troublesome rule.

Unfortunately, studies have not been done on the relationships among learning indicators and outcomes. Most of the evaluation studies have simply compared LISP-tutored students with classroom or human-tutored students on a standard achievement test administered at the end of the course. However, one study did investigate individual differences in acquisition and retention of individual productions over a series of ten lesson sessions (reported by Anderson, in press). In this analysis, each production was scored for the number of times it was used incorrectly in problem solving, separately for each session. A series of factor analyses were performed on these data to determine whether production factors would emerge. For example, it could be that productions associated with one *kind* of learning (such as learning to trace functions or planning) would form a factor separate from some other *kind* of learning (such as learning to select functions or coding). Or lesson-specific factors could have emerged. In fact, Anderson found evidence for two broad factors: an *acquisition factor* captured individual differences in speed of production acquisition and a *retention factor* captured individual differences in the likelihood that acquired productions were retained in a later session.

Applying the Taxonomy Consider first how we might classify the LISP tutor. Students spend most of their time learning specific production rules and skills and are continually tested for their ability to apply them in writing LISP functions. Every student action can be viewed as a test response because the system is interpreting that response as an indication of whether the student knows a particular production rule. Thus, learning and testing activities in the LISP tutor are almost completely integrated.

Although students are learning *skills*, insofar as writing functions is a multiple operator selection task, the LISP tutor is testing for students' knowledge of the *rules* underlying those skills. But this merely reflects the fact that skills in the LISP tutor are defined precisely in terms of their constituent rules. Interestingly, the fact that the LISP tutor can represent students' skill without directly evaluating that skill (in other words, the system never evaluates whether the function works, per se) is evidence against the taxonomy's supposition of skill as a separate knowledge type. However, this presumes a rule-level understanding of skill. In domains for which such a detailed understanding is not yet available (most domains imaginable at this time), skill probably ought to be considered a functionally distinct category even if only for pragmatic reasons.

The instructional environment is one in which students learn initially through brief instruction (a pamphlet or a textbook), but then go on to compile and refine that knowledge by engaging in extended problem solving. Figure 4-3b summarizes our assessment of what learning skills are being exercised and tested in the LISP tutor.

Note that in addition to indicating that students are learning declarative knowledge by instruction, and procedural knowledge by compiling and practicing it, we have indicated other learning products and sources. The other products are the general rules and skills probably being taught by the LISP tutor, even though that is not a goal for the tutor. The other sources have to do with the LISP tutor's capability of delivering context-sensitive tutorial advice, and through its coaching capabilities can readily change the nature of the instructional environment. On one occasion it might correct a student's attempt through direct instruction, but then it might later suggest an analogy to a student, or provide examples of a concept.

Now consider the testing comprehensiveness issue. As can be seen in Figure 4-3b, we consider all of the LISP tutor's testing to be for rule knowledge either in the compilation or the drill and practice environments. (We could also consider automatic skills to be tested, but that would require a rather detailed analysis of the LISP tutor's entire production collection of how big, compiled productions subsume their smaller precursors.) Note that first, as with BIP, students' success at propositional learning and their ability to acquire general rules and skills is not tested. This situation could be remedied with the insertion of sentence verification and transfer-of-training tests. But a more intriguing suggestion from the standpoint of research arises from the LISP tutor's multifaceted coaching capability, which offers various kinds of tutorial remediation, to greatly expand the range of learning events that may be investigated. For example, it would be possible (and interesting) to keep track of production strength modification separately for each of the various instructional environments. That is, one could trace the growth in rule indicators over time as a function of whether those rules were taught (or remediated) with instructional advice, analogies, examples, and so on. One could ask, for example, whether instruction using analogies results in greater subsequent ability to use the rule(s) so instructed.

In summary, because of the way in which it models students' knowledge as production rules, and carefully controls the learning environment, the LISP

tutor is ideally suited for measuring learning skills such as the rate at which productions are composed, or the probability of compiling a sequence of productions as a function of exposure to that sequence. Augmented with the additional tests and performance records suggested by the application of the taxonomy, the LISP tutor could serve as an excellent research tool for investigating the time course of learning and individual differences therein.

Smithtown: Discovery World for Economic Principles

General System Description Unlike the other two systems, Smithtown's main goal is to enhance students' general problem-solving and inductive-learning skills. It does this in the substantive context of microeconomics in teaching the laws of supply and demand (Shute and Glaser, in press). Smithtown is highly interactive. Students pose questions and conduct experiments within the computer environment, testing and enriching their knowledge of functional relationships by manipulating various economic factors.

As a discovery environment, Smithtown is quite different from BIP and the LISP tutor in that there is no fixed curriculum. The student—not the system—generates problems and hypotheses. After generating a hypothesis such as "Does increasing the price of coffee affect the supply or demand of tea?" the student tests it by executing a series of actions, such as changing the values of two variables and observing the bivariate plot. This series of actions, or *behaviors,* for creating, executing, and following up a given experiment, define a *student solution.*

Despite having no curriculum, Smithtown does have the instructional goal of teaching general problem-solving rules and skills (called *good critics*) such as "collect baseline data before altering a variable" or "generalize a concept across two unrelated goods." Instead of a curriculum guiding instructional decisions, Smithtown relies on a process of constantly monitoring student actions, looking for evidence of good and poor behavior, then coaching students to become more effective problem solvers. The system keeps a detailed history list of all student actions, grouping them into (that is, interpreting them as) behaviors and solutions. Smithtown diagnoses solution quality in two ways. It looks for overt errors by comparing student solutions with its *buggy critics,* which are sets of actions (or nonactions) that constitute nonoptimal behaviors (for example, "fail to record relevant data in the online notebook"). It also compares student solutions with its own *good critics* (expert solutions). Discrepancies between the two are collected into a list of potential problem areas and passed on to the coach for possible remediation. To illustrate, if the student had failed to enter data into the online notebook for several time frames and had made some changes to variables, the system would recognize this as a deficient pattern and prompt the student to start using the notebook more consistently.

Smithtown's *student model* is based on two statistics: (a) the number of times the student demonstrates a buggy critic (errors of commission), and (b) the ratio of the number of times the student uses a good critic over the number of times it was applicable (errors of omission). Coaching is based on the heuristic of first advising about buggy behaviors, then advising on any blatant errors of omission. Advice is always given in the context of a particular experiment, so, like the LISP tutor, it is context sensitive. For example, the coach might say, "You haven't graphed any data yet and I think you should try it out. This is often a good way of viewing data. It lets you plot variables together and some surprising relationships may become apparent." However, the coach is fairly unobtrusive: after advice is given, there is no further coaching for some time.

Smithtown also knows about variable relationships that constitute economics principles, such as "Price is inversely related to quantity demanded." If a student uses the system's hypothesis menu and states this relationship (for example, "As price increases, quantity demanded decreases"), the student is congratulated and told the name of the law just discovered ("Congratulations! You have just discovered what economists refer to as the *Law of Demand*").

Learning Indicators Shute, Glaser, and Raghavan (Chapter 8 of this volume) conducted an extensive evaluation of differences among students in what they learned and how they interacted with Smithtown. Two data sources were used: a list of all student actions and a set of verbal protocols in which students justified their actions and predicted outcomes of the actions.

Table 4-4 shows a set of twenty-nine learning indicators constructed for analyzing individuals' performance. Indicators are clustered into three general behavior categories: (a) *activity exploratory level* (indicators relating to activity level and exploratory behaviors), (b) *data management level* (indicators for data recording, efficient tool use, and use of evidence), and (c) *thinking and planning level* (indicators for consistent behaviors, effective generalization, and effective experimental behaviors).

Shute et al.'s sample ($N = 10$) was too small to analyze formally, but the indicators were examined to determine which ones discriminated successful from unsuccessful learners. Two subjects, one who performed poorly on the pretest but well on the posttest (a successful learner) and one who who did poorly on both tests (an unsuccessful learner) were selected for more careful scrutiny.

The two subjects differed mostly on indicators of thinking and planning skills, in other words, effective experimental behaviors. In particular, the better subject collected and organized data from a more theory-driven perspective, which contrasted with a more superficial and less theory-driven approach used by the poorer subject. The better subject generalized concepts across multiple markets (which the poorer subject did not do), engaged in more investigations

Table 4-4 Learning Indicators from Smithtown, the Economics Tutor

Activity and exploratory level skills

I. Activity level
   1. Total number of actions
   2. Total number of experiments
   3. Number of changes to the price of the goods

II. Exploratory behaviors (counts; i.e., number of . . .)
   4. Markets investigated
   5. Independent variables changed
   6. Computer-adjusted prices
   7. Times market sales information was viewed
   8. Baseline data observations of market in equilibrium

Data-management level skills

III. Data recording
   9. Total number of notebook entries
   10. Number of baseline data entries of market in equilibrium
   11. Entry of changed independent variables

IV. Efficient tool usage (ratios of number of effective uses over number of uses)
   12. Number of relevant notebook entries ÷ total number of notebook entries
   13. Number of correct uses of table package ÷ number of times table used
   14. Number of correct uses of graph package ÷ number of times graph used

V. Use of evidence
   15. Number of specific predictions ÷ number of general hypotheses
   16. Number of correct hypotheses ÷ number of hypotheses

Thinking and planning level skills

VI. Consistent behaviors (counts; i.e., number of . . .)
   17. Notebook entries of planning menu items
   18. Notebook entries of planning menu items ÷ planning opportunities
   19. Number of times variables were changed that had been specified beforehand in the planning menu

VII. Effective generalization (event counts; i.e., number of times . . .)
   20. An experiment was replicated
   21. A concept was generalized across unrelated goods
   22. A concept was generalized across related goods
   23. The student had sufficient data for a generalization

VIII. Effective experimental behaviors (event counts; i.e., number of times . . .)
   24. A change to an independent variable was sufficiently large
   25. One of the experimental frames was selected
   26. The prediction menu was used to specify an event outcome
   27. A variable was changed (per experiment)
   28. An action was taken (per experiment)
   29. An economic concept was learned (per session)

poorer subject. The better subject also made large changes to variables so that any repercussions could be detected. This contrasted with typically small changes made by the poorer subject, who justified these choices by claiming they were more "realistic." Replicating experiments to test the validity of results is an important scientific behavior and similar to BIP's indicator 15. The better subject conscientiously replicated experiments whereas the poorer subject did not. One other indicator, data management skills, distinguished the two subjects. The better subject recorded more notebook entries, and the ones that were recorded consistently included relevant variables from the planning menu. The poorer subject used the notebook sporadically and often failed to record important information.

Applying the Taxonomy Again, we first consider the classification of Smithtown. *Knowledge types* taught are primarily general skills (that is, effective inquiry strategies for a new domain), domain-specific skills pertaining to economics knowledge, and domain-specific mental models of the functional relationships among microeconomic factors. Students also are presumed to acquire some declarative knowledge and rules about economics while interacting with the environment. The *instructional environment* is a discovery microworld, and thus most of the learning that occurs results from students inducing knowledge and skills through observation and discovery, then perhaps compiling those skills by practicing them in the conduct of experiments. There is tutorial assistance if a student is judged to be floundering in discovery mode, however; we indicate this in Figure 4-3c as learning propositions and skills by direct instruction. Figure 4-3c shows that in overall emphasis, Smithtown is quite distinct in both goals and approach from BIP and the LISP tutor.

Regarding the issue of testing comprehensiveness in Smithtown, we consider two kinds of tests: (a) the on-line indicators used by the system in diagnosis, and (b) the separate posttest that measured economics knowledge gained during the tutorial. For the purpose of filling out Figure 4-3c, we considered half the total testing to be on-line and the other half to be the posttest; the striped bars are marked as to the testing source. Figure 4-3c shows that as in the LISP tutor, the on-line indicators primarily reflect rule and skill knowledge, but in Smithtown the testing context is the discovery environment. Another key difference is that the rule and skill knowledge is not related to the economics domain, but rather to the subject's ability to manipulate the environment and use its tools to test hypotheses. The posttest did tap domain knowledge. One part of the posttest battery was a multiple choice test that measured declarative knowledge. A second part was a scenarios test that had subjects reason through various economics scenarios. The scenarios test illustrates a means for assessing mental models: it was designed to get at students' ability to run mental simulations of complex economics scenarios (see Shute and Glaser, in press, for a detailed discussion of the test).

Figure 4-3c suggests that perhaps the greatest mismatch between learning skills that were exercised and those that were tested occurs in the general rule and skill cells. A shortcoming of the Smithtown evaluation is that one of its stated primary goals is to help students become more effective in conducting experiments in a microworld environment, acquiring general skills as a result of their investigations. But this instructional goal was measured only indirectly on the posttest, which relied on declarative tests of economics knowledge. A more direct assessment of the degree to which the stated goals could be reached would require a transfer of skills in a system structured similarly to Smithtown but containing different domain knowledge. (Interestingly, there is such a system, but the transfer experiment has not yet been conducted.) Truly general inquiry skills developed in Smithtown would presumably transfer to the new environment.

Another smaller mismatch is that declarative knowledge of basic economic principles was tested at posttest, but not while students were interacting with the tutor. It seems reasonable from both a research standpoint and the standpoint of enhancing the student model to integrate declarative knowledge tests with tutoring.

A major factor missing here and throughout our discussion of the three tutors is the style dimension. Inspection of Table 4-4 shows that the set of indicators Smithtown collects and monitors are really not direct indicators of learning skill per se, but rather are style indicators in the sense that they reveal how an individual organizes his or her learning environment. From this perspective, key questions addressed in the Shute et al. analysis had to do with style interrelationships (the question of dimensionality of style) and the relationship between style and learning outcome (the question of validity). In one sense, this is exactly the study needed to understand learning skills in the most natural, ecologically valid context. It is also a preliminary question to one of the goals we are pushing for here: to be able to assess basic learning skills, controlling for learning style. Smithtown may be best suited for analysis of the style issue. But before style variables are better understood, more structured environments such as BIP and the LISP tutor, which by forcibly directing learning activities designate a less important role for individual variability in learning style, may be more conducive to research on basic learning skills.

LEARNING INDICATORS FOR VALIDATION STUDIES

To this point, we have discussed how the taxonomy might be applied so as to enable a more thorough evaluation of student learning skills and outcomes. The applications discussed above might have the flavor of suggestions for improving the tutors. That is not the intention. We see the main function of the taxonomy as primarily a research one. By more thoroughly examining what students learn in instruction, it should be possible to conduct more refined studies on individual differences in learning. Snow et al. generated and analyzed a set of learning indicators, Anderson did a similar analysis, and a similar analysis is underway for Smithtown. Our claim is that the taxonomy should suggest additional ways in which to record learning skills, and this should result in a psychologically rich and principled set of additional learning indicators. Each cell in the full four-dimensional taxonomy defines a proposed learning skill. An important next question, open to empirical investigation, concerns the true reduced-space dimensionality of learning skills. From an individual differences perspective, how many learning *abilities* must we posit, and at what level of detail, to characterize *skill* differences among learners over all taxonomy cell tasks?

There is also a second, related application. The taxonomy should help us develop learning indicators for instructional programs that can serve as criteria against which other individual difference measures, such as aptitude and basic abilities tests, might be validated. That is, our taxonomy-derived indicators can serve as supplements or even replacements for the conventional criteria of post-course achievement tests, course grade point average, on-the-job performance tests, and supervisor or teacher ratings in the conduct of construct validation studies. Indeed, it was this goal of creating more extensive criteria against which new aptitude tests might be validated that led us into the taxonomy project in the first place.

Learning Abilities Measurement Program (LAMP)

Over the past several years, the Air Force has supported a basic program of research designed to explore the possibility of using contemporary cognitive theory as the basis for a new system of ability measurement (Kyllonen, 1986; Kyllonen and Christal, in press). Currently, the Air Force, as well as the other services, selects and assigns applicants at least partly on the basis of their performance on a conventional aptitude battery, which includes tests of reading comprehension, arithmetic reasoning, numerical operations, and so forth. The goal of the Learning Abilities Measurement Program (LAMP) is to provide the research base that might lead to supplementing or even replacing those conventional tests with new measures more closely aligned with an information-processing perspective.

What might these new tests be? The project has thus far investigated measures of working-memory capacity, information-processing speed, breadth and depth of declarative knowledge, availability of strategic knowledge, and other such abilities. It would go beyond the scope of this chapter to review the project's research (see Kyllonen, 1986; Kyllonen and Christal, in press, for current reviews), but the prototypical study investigates the relationship between various kinds of cognitive measures (such as working-memory capacity) and learning-outcome measures (list recall) under various instructional conditions (such as variations in study time).

A major focus of the research is examining the relationships between ability measures and learning outcomes. But the range of learning outcomes investigated thus far, not only on our project but on others as well, has been quite limited in two ways. First, the range of learning skills examined has been rather narrow: this is especially apparent given the breadth of potential learning skills suggested by the taxonomy. But second, and perhaps even more important, learning tasks employed have not been truly representative of real-world learning activities. Tasks tend to be short-term laboratory tasks, which afford more control, but also leave bigger validity gaps with the kind of operational learning to which we eventually wish to generalize. This inhibits the transition of research to application, insofar as generalization from narrow laboratory tasks to real-world learning tasks is tenuous. And as Greeno (1980) has argued, use of ecologically valid learning tasks is defensible from the standpoint of leading to better basic research as well.

Thus, for both applied and theoretical reasons, a decision was made recently to expand the range of learning criteria employed. A recently completed laboratory at Lackland Air Force Base accommodates thirty work stations capable of administering intelligent computerized instruction like that reviewed previously. Intelligent tutoring systems in the domains of computer programming, electronic troubleshooting, and flight engineering have been developed or are currently underway. Over the next several years we will investigate learning on these tutors and conduct studies that examine the relationship between basic cognitive abilities and various learning skills and outcomes. We expect the taxonomy as described here to assist us in developing learning indicators for the tutorial environments.

Applying the Taxonomy: A Practical Guide

Thus, we are employing a two-pronged approach in generating learning skill indicators for LAMP validation studies. We design instructional programs capable of producing rich traces of learner activities, then we intend to analyze and categorize those activities so as produce psychologically meaningful learning indicators. Tables 4-5 and 4-6 present the general outline for our approach. Note that we have written the design and analysis steps in such a way as to be broadly useful. Although our application is in the design and (especially) analysis of intelligent tutoring systems, the steps suggested could be adapted to any kind of instructional system, computerized or even classroom.

Table 4–5 Applications of the Taxonomy: Instructional System Design Steps

1. Determine desired *knowledge outcomes*
   a. State the instructional goals (e.g., acquisition of a mental model, a set of propositions, a set of skills).
   b. Specify the particular facts, skills, or mental models to be taught.
   c. Determine tests to be used for assessing particular knowledge outcomes (Table 4-2).
2. Determine *environment* for achieving knowledge outcomes
   a. Consider the kind of learning strategy desirable to invoke (Table 4-1).
   b. Consider alternative means for achieving knowledge outcome (could be used as a remediation strategy, or simply as a variation to avoid instructional monotony).
   c. Record student learning success with respect to the knowledge-outcome-by-instructional-environment matrix. This allows more precise statements of the effectiveness of the instruction.
3. Consider *learning style* issues
   a. Consider whether to encourage particular types (styles) of interaction.
   b. If learning style is left free, make provisions to record the manner in which the student interacts with the instructional environment (for suggestions see Tables 4-3 and 4-4). This also allows more precise statements of the effectiveness of the instruction.
   c. If particular learning styles are encouraged through feedback and suggestions, consider varying the kinds of styles encouraged so as to allow experimental comparisons of the relative effectiveness of various styles.

SUMMARY AND DISCUSSION

We have presented a taxonomy of learning based on previous research and on contemporary cognitive theory. We have also proposed how the taxonomy can be applied to generate indicators of what a student in an instructional situation is learning, and how well he or she is learning it. But just how well does our proposed taxonomy-indicator system work?

Consider four major uses for the system (these and a fifth research application are listed in Table 4-7). First, the taxonomy can suggest what kinds of skills are being exercised and tested in an instructional setting. In this capacity the taxonomy serves in much the same way Bloom's or Gagné's taxonomies do. The advantage of our proposal is that it is more closely tied to current cognitive theory, which we hope will enable us to apply the system more easily in analyzing learning in somewhat naturalistic instructional settings. A second use for the system concerns primarily the environment dimension. The specification of multiple instructional environments provides a way to think about a range of means for achieving particular knowledge outcomes. If an instructor's

Table 4–6 Applications of the Taxonomy: Learning Task Analysis Steps

1. Determine the *knowledge outcome* goals for the instruction
   a. Determine the nature of the stated instructional goals (e.g., acquisition of a mental model, a set of propositions, a set of skills).
   b. Determine what kinds of tests are embedded within the instruction (consulting Table 4-2).
   c. Determine the match between the tests used and the knowledge outcomes intended (as in Figure 4-3).
2. Determine the nature of the *instructional environment*
   a. For every instructional exchange (every student-instructor interaction episode), consider what learning strategy is invoked (consulting Table 4-1) during the exchange. Generate learning activities profiles for the entire instructional program (as in Figure 4-3).
   b. Organize records of student learning success with respect to the knowledge-outcome-by-instructional-environment (KO $\times$ IE) matrix. That is, devise a means for assigning each student a separate learning success score for each cell in the KO $\times$ IE matrix. Scores would be based on tests following particular instructional exchanges.
3. Consider *learning style* issues
   a. Consider whether particular types (styles) of interaction are encouraged.
   b. If learning style is left free, and there is between-student style variability, but no within-student style variability, then separate students by style before conducting any analyses of the KO $\times$ IE matrix.
   c. If learning style is left free, and there is within-student style variability (e.g., students engage in holistic processing some times, serial processing at others), create separate KO $\times$ IE profiles for the various style orientations.
4. Considerations for transfer studies
   a. Degree of transfer should be a function of the similarity of the learning activities profiles for two learning tasks.
   b. Similarity is computed over the KO $\times$ IE matrices (possibly for separate styles) and *domain*.
5. Considerations for optimizing or predicting global outcomes
   a. Expected global outcome for a particular student will depend on the match between the student's personal learning-skill profile and the learning skills the instruction exercises (the learning activies profile, Figure 4-3).
   b. Optimizing global outcomes for a particular student can be seen as a linear programming problem. Instruction should maximize exercising the student's strongest skills subject to the cost (e.g., in time) for exercising those skills.

goal is to teach a mental model of some system, the instructor can simply instruct it, use an analogy, have the student discover the model through observation of the system, or employ another instructional approach. A third use for the system is to make predictions about transfer relations among learning experiences. We would predict that the closer, taxonomically, two learning

Table 4–7 Applications of the Taxonomy: What It Can Be Used For

Instructional system evaluators (teachers and administrators)
1. Facilitates analysis of the kinds of learning skills that are being exercised and tested in an instructional setting (see Figure 4-3)

Instructional system designers
2. Suggests a range of possible instructional environments for achieving particular knowledge outcomes (see Table 4-1/Figure 4-1)
3. Specifies techniques (tests) for probing a wide range of knowledge and learning skill outcomes (see Table 4-2)

Cognitive researchers
4. Suggests predictions about transfer relations among learning experiences (see Figure 4-1/Table 4-6)
5. Suggests indicators (dependent variables) of what and how well a student is learning (see Figure 4-3/Tables 4-2, 4-6)

situations are, the more likely that whatever is learned in one will transfer to the other. Of course this is an open empirical question. A benefit of the taxonomy is that it suggests a straightforward research program for addressing this kind of question.

While all three of these applications may be useful, we believe that the most important role of the taxonomy is in establishing the means for probing a much wider range of knowledge and learning-skill outcomes. This capability is obviously important for research purposes, but it also is important for evaluating educational systems. Consider a general problem in evaluating innovative educational programs (discussed by Nickerson, Perkins, and Smith, 1985). Over the years many such programs, such as ones for teaching creative thinking or ones for teaching general thinking skills, have been developed. All too often, casual observation suggests that such programs are having desirable effects on students, but such effects do not show up under the scrutiny of carefully conducted evaluation studies. Creators of such programs typically complain that the scientific model of evaluation is inappropriate because the true gains students experience are somehow missed. One role for the taxonomy might be to suggest how additional learning outcomes and skills can be assessed in order to enable a more thorough evaluation.

Even among the three instructional programs we reviewed here, a rather conservative approach to assessing the impact of the tutoring system was taken. To some extent, the LISP tutor, BIP, and Smithtown all depended on standard achievement outcome tests as a means for their validation. While it is important to establish that these tutors do affect overall achievement, it is not sufficient. While interacting with a tutor, or in any instructional environment, students can be learning many different things. A major role for the taxonomy is to suggest a richer testing system for evaluating a broader range of student outcomes.

Finally, the taxonomy-indicator system should facilitate pursuit of both applied and basic research questions. Our major practical application for the taxonomy is to have it assist in the specification of variables that indicate what and how well a subject is learning as the subject interacts with a tutor over a lengthy series of lessons. These variables then will serve as criteria against which newly developed measures of cognitive ability will be validated. In addition, a wide range of basic research issues is opened up. Are the different knowledge types affected by the same variables? Are fast propositional learners also fast production-rule learners? Are there interactions between knowledge type and the instructional environment? Are individual differences in learning more dependent on the knowledge type or the environment? Our research programs are only at the very beginning stages in addressing these kinds of fundamental questions about the nature of learning and individual differences therein.

ACKNOWLEDGMENTS

Support was provided by the Air Force Human Resources Laboratory and the Air Force Office of Scientific Research, through Universal Energy Systems, under Contract No. F41689-84-D-0002/58420360, Subcontract No. S-744-031-001. We thank Raymond Christal, Wesley Regian, Dan Woltz, William Alley, Stephanie Rude, Bruce Britton, and Shawn Glynn for their valuable comments on the taxonomy.

REFERENCES

Allison, R. B., Jr. (1960). *Learning parameters and human abilities.* Unpublished doctoral dissertation, Educational Testing Service and Princeton University.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, Mass.: Harvard University Press.

Anderson, J. R. (1987a). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review,* 94:192–210.

Anderson, J. R. (1987b). Production systems, learning, and tutoring. In D. Klahr, P. Langley, and R. Neches (Eds.), *Production system models of learning and development.* Cambridge, Mass.: MIT Press.

Anderson, J. R. (In press). Analysis of student performance with the LISP tutor. In N. Frederiksen (Ed.), *Diagnostic monitoring of skill and knowledge acquisition.* Hillsdale, N.J.: Erlbaum.

Anderson, J. R., Boyle, F., and Reiser, B. (1985). Intelligent tutoring systems. *Science,* 228:456–462.

Baron, J. Badgio, P. C., and Gaskins, I. W. (1986). Cognitive style and its improvement:

A normative approach. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence,* Vol. 3. Hillsdale, N.J.: Erlbaum.

Barr, A., Beard, M., and Atkinson, R. C. (1976). The computer as a tutorial laboratory: The Stanford BIP project. *International Journal of Man-Machine Studies,* 8:567–596.

Bloom, B. S. (1956). *Taxonomy of educational objectives: Cognitive domain (Handbook I).* New York: McKay.

Campione, J. C., Brown, A. L., and Bryant, N. R. (1985). Individual differences in learning and memory. In R. J. Sternberg (Ed.), *Human abilities: An information processing approach.* New York: W. H. Freeman and Company.

Carbonell, J. G., Michalski, R. S., and Mitchell, T. M. (1983). An overview of machine learning. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach.* Los Altos, Calif.: Morgan Kaufmann.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Chi, M. T. H., Feltovich, P. J., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science,* 5:121–152.

Cronbach, L. J., and Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington.

Ennis, R. H. (1986). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron and R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* New York: W. H. Freeman and Company.

Estes, W. K. (1982). Learning, memory, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence.* Cambridge: Cambridge University Press.

Fitts, P. M. (1963). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* New York: Academic Press.

Fleishman, E. A., and Quaintance, M. K. (1984). *Taxonomies of human performance: The description of human tasks.* Orlando, Fla.: Academic Press.

Fredericksen, N., and Pine, J. (In press). Assessment of learning in science and mathematics. *Report on indicators of quality of education in science and math.* Washington, D.C.: National Academy of Sciences/National Research Council.

Gagné, R. M. (1965). Problem solving. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press.

Gagné, R. M. (1985). *The conditions of learning and theory of instruction.* New York: Holt, Rinehart & Winston.

Gagné, R. M., and Briggs, L. J. (1979). *Principles of instructional design,* 2d ed. New York: Holt, Rinehart & Winston.

Gitomer, D. (1984). A cognitive analysis of a complex troubleshooting task. Unpublished doctoral dissertation. University of Pittsburgh.

Glaser, R. (1967). Some implications of previous work on learning and individual differences. In R. M. Gagné (Ed.), *Learning and individual differences.* Columbus, Ohio: Merrill.

Glaser, R., Lesgold, A., and Lajoie, S. (In press). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. Glover, J. C. Conoley, and J. C. Witt (Eds.), *The influence of cognitive psychology on testing, Buros/Nebraska symposium on testing,* Vol. 3. Hillsdale, N.J.: Erlbaum.

Glynn, S. M., Britton, B. K., Semrud-Clikeman, and Muth, D. K. (In press). Analogical reasoning and problem solving in science textbooks. In J. A. Glover, R. R. Ronning, and C. R. Reynolds (Eds.), *Handbook of creativity: Assessment, research, and theory.* New York: Plenum.

Goodenough, D. R. (1976). The role of individual differences in field dependence as a factor in learning and memory. *Psychological Bulletin,* 83:675–694

Greeno, J. G. (1980). Psychology of learning, 1960–1980: One participant's observations. *American Psychologist,* 35:713–728.

Hayes, J. R., and Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg and E. R. Steinberg (Eds.), *Cognitive processes in writing.* Hillsdale, N.J.: Erlbaum.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence,* 3:105–120.

Jensen, A. R. (1967). Varieties of individual differences in learning. In R. M. Gagne (Ed.), *Learning and individual differences.* Columbus, Ohio: Merrill.

Kyllonen, P. C. (1986). Theory based cognitive assessment. In J. Zeidner (Ed.), *Human productivity enhancement,* Vol. 2: *Acquisition and development of personnel.* New York: Praeger.

Kyllonen, P. C., and Christal, R. E. (In press). Cognitive modeling of learning abilities: A status report of LAMP. In R. Dillon and J. W. Pellegrino, *Advances in testing and training.* New York: Academic Press.

Langley, P. (1986). Editorial: On machine learning. *Machine Learning,* 1:5–10.

Linn, M. C., and Kyllonen, P. C. (1981). The field dependence-independence construct: Some, one, or none. *Journal of Educational Psychology,* 73:261–273.

Linville, P. A. (1982). Affective consequences of complexity regarding the self and others. In M. S. Clark and S. T. Fiske (Eds.), *Affect and cognition: The seventeenth annual Carnegie symposium on cognition.* Hillsdale, N.J.: Erlbaum.

Malmi, R. A., Underwood, B. J., and Carroll, J. B. (1979). The interrelationships among some associative learning tasks. *Bulletin of the Psychonomic Society,* 13:121–123.

Marshalek, B., Lohman, D. F., and Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence,* 7:107–127.

Martin, D. W. (1986). Turning research art into science. *Contemporary Psychology,* 31:211–212.

Meichenbaum, D. (1977). *Cognitive behavior modification: An integrative approach.* New York: Plenum.

Melton, A. W. (1964). The taxonomy of human learning: Overview. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press.

Melton, A. W. (1967). Individual differences and theoretical variables: General comments on the conference. In R. M. Gagne (Ed.), *Learning and individual differences.* Columbus, Ohio: Merrill.

Messick, S. (1986). Cognitive style. In R. E. Snow and M. J. Farr (Eds.), *Aptitude, learning, and instruction,* Vol. 3: *Conative and affective process analysis.* Hillsdale, N.J.: Erlbaum.

Michalski, R. S. (1986). Understanding the nature of learning: Issues and research directions. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach,* Vol. II. Los Altos, Calif.: Morgan Kaufmann.

Murnane, R. J., and Raizen, S. A. (1988). *Improving indicators of the quality of science and mathematics education in grades K–12.* Washington, D.C.: National Academy Press.
Nickerson, R. S., Perkins, D. N., and Smith, E. E. (1985). *The teaching of thinking.* Hillsdale, N.J.: Erlbaum.
Pask, G., and Scott, B. C. E. (1972). Learning strategies and individual competence. *International Journal of Man-Machine Studies,* 4:217–253.
Perrig, W., and Kintsch, W. (1984). *Propositional and nonpropositional representations of text* (Tech. Rep. No. 131). Boulder, Colo.: Institute of Cognitive Science, University of Colorado.
Reiser, B. J., Anderson, J. R., and Farrell, R. G. (1985). *Dynamic student modeling in an intelligent tutor for LISP programming.* Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Vol. 1. Los Altos, Calif.: Morgan Kaufmann.
Rosenbloom, R. S., and Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.) *Machine learning: An artificial intelligence approach,* Vol. II. Los Altos, Calif.: Morgan Kaufmann.
Rumelhart, D. E., and Norman, D. A. (1981). Analogical processes in learning. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, N.J.: Erlbaum.
Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review,* 84:1–66.
Schoenfeld, A. H. (1985). *Mathematical problem solving.* New York: Academic Press.
Self, J. (1986). The application of machine learning to student modelling. *Instructional Science,* 14:327–338.
Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review,* 84:127–190.
Shute, V. J., and Glaser, R. (In press). An intelligent tutoring system for exploring principles of economics. In R. E. Snow and D. Wiley (Eds.), *Straight thinking.* San Francisco: Jossey Bass.
Snow, R. E., Kyllonen, P. C., and Marshalek, B. (1984). The topography of learning and ability correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence,* Vol. 2. Hillsdale, N.J.: Erlbaum.
Snow, R. E., Wescourt K., and Collins, J. (1986). *Individual differences in aptitude and learning from interactive computer-based instruction* (Tech. Rep. No. 10). Stanford, Calif.: Standford University School of Education.
Soloway, E., and Littman, D. (1986). *Evaluating ITSs: The cognitive science perspective.* Proceedings of the Research Planning Forum for Intelligent Tutoring Systems. San Antonio, Tex.: Air Force Human Resources Laboratory.
Spelke, E., Hirst, W., and Neisser, U. (1976). Skills of divided attention. *Cognition,* 4:215–230.
Stake, R. E. (1961). Learning parameters, aptitudes, and achievement. *Psychometric monographs,* No. 9.
Thurstone, L. L. (1938). Primary mental abilities. *Psychometric monographs,* No. 1.
Tversky, A., and Hutchins, J. W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review,* 93:3–22.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist,* 30:128–134.
Underwood, B. J., Boruch, R. F., and Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General,* 107:393–419.
Wescourt, K. T., Beard, M., Gould, L., and Barr A. (October, 1977). *Knowledge-based CAI: CINS for individualized curriculum sequencing* (Tech. Rep. No. 290). Stanford, Calif.: Institute for Mathematical Studies in the Social Sciences, Stanford University.
Wickens, C. D., Sandry, D., and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output: Testing a model of complex task performance. *Human Factors,* 25:227–248.
Winograd, T. (1975). Frame representations and the declarative-procedural controversy. In D. G. Bobrow and A. Collins (Eds.), *Representation and understanding.* New York: Academic Press.
Woodrow, H. (1946). The ability to learn. *Psychological Review,* 53:147–158.
Yazdani, M. (1986). Intelligent tutoring systems survey. *Artificial Intelligence Review,* 1:43–52.