DTIC FILE COPY

(2)

AFHRL-TP-87-39

AIR FORCE

HUMAN RESOURCES

AD-A190 669

TAXONOMY OF LEARNING SKILLS

Patrick C. Kyllonen

Institute for Behavioral Research
University of Georgia
Athens, Georgia 30602

Valerie J. Shute

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

April 1988
Interim Technical Paper for Period February 1986 - 1987

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS: AD-A190669
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFHRL-TP-87-39
6a. NAME OF PERFORMING ORGANIZATION: Universal Energy Systems; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Manpower and Personnel Division
6c. ADDRESS (City, State, and ZIP Code): 4401 Dayton-Xenia Road, Dayton, Ohio 45432	7b. ADDRESS (City, State, and ZIP Code): Air Force Human Resources Laboratory, Brooks Air Force Base, Texas 78235-5601
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Air Force Human Resources Laboratory; 8b. OFFICE SYMBOL (If applicable): HQ AFHRL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F41689-84-D-0002/58420360 S-744-031-001
8c. ADDRESS (City, State, and ZIP Code): Brooks Air Force Base, Texas 78235-5601	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2313; TASK NO T1; WORK UNIT ACCESSION NO. 33

11. TITLE (Include Security Classification)
Taxonomy of Learining Skills

12. PERSONAL AUTHOR(S)
Kyllonen, P.C.; Shute, V.J.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM Feb 86 TO Feb 87	April 1988	66

16. SUPPLEMENTARY NOTATION
To appear in P.L. Ackerman, R.J. Sternberg, & R. Glaser (Eds.), Learning and individual differences. San Francisco: Freeman.

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	08	
05	09	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
cognition; cognitive ability; computerized testing; individual differences; learning; learning ability

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Questions concerning individual differences in learning ability may be more precisely addressed in light of an agreed-upon taxonomy of learning skills. In this paper we review a variety of attempts to propose learning taxonomies and point out some of their limitations. We then propose a taxonomy consisting of four dimensions: (a) learning environment (or learning process), (b) resulting knowledge type, (c) domain, and (d) learning style. The first three dimensions specify a learning task, and the fourth is a modifiable learner characteristic. We apply the taxonomy to the analysis of three computerized instructional systems in an attempt to answer the question of (a) what learning skills the systems currently exercise, (b) what skills the systems test, and (c) what other skills might fruitfully be tested. We conclude with a discussion of how the taxonomy might be used to guide research that attempts to validate new aptitude tests against performance in complex learning situations.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Nancy J. Allin, Chief, STINFO Office	22b. TELEPHONE (Include Area Code): (512) 536-3877; 22c. OFFICE SYMBOL: AFHRL/TSR

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

AFHRL Technical Paper 87-39

April 1988

TAXONOMY OF LEARNING SKILLS

Patrick C. Kyllonen

Institute for Behavioral Research
University of Georgia
Athens, Georgia 30602

Valeria J. Shute

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Accesion For	
NTIS CRA&I	✓
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

Reviewed by

Dan J. Woltz
Learning Abilities Research Function

Submitted for publication by

Lloyd D. Burtch
Chief, Cognitive Assessment Branch

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

The Air Force's Learning Abilities Measurement Program (LAMP) conducts basic research on the nature of human learning abilities, with the ultimate goal of contributing to an improved personnel selection and classification system. To date, studies in the program have investigated the relationship between aptitude measures and performance on simple learning tasks. One limitation to these studies is that it may be inappropriate to generalize results obtained to an operational setting. Thus, future efforts will validate the aptitude tests against more complex learning such as computer programming, electronic troubleshooting, flight engineering, and air traffic control.

Before the newer effort is underway, it is critical to give serious attention to the question of how learning might be measured in more complex environments. In this paper, we demonstrate how learning indicators may be derived from a taxonomy of learning to ensure that a wide range of learning outcomes will be assessed during instruction. The paper first reviews existing taxonomies, and points out their limitations. A taxonomy is then proposed based on a synthesis of current thought regarding the forms of knowledge, the types of learning activities, the importance of the domain, and the effects of the learner's style. The taxonomy is applied to analyze some computerized instructional programs that attempt to measure student learning, and show how the programs might be improved by measuring a broader variety of learning outcomes. The paper concludes by speculating about how the taxonomy aids consideration of a broad variety of questions concerning the relationships between basic cognitive skills and learning outcomes, and the relationships among different kinds of learning experiences.

PREFACE

Development of this paper was supported by the Air Force Learning Abilities Measurement Program (LAMP), a multi-year program of basic research conducted at the Air Force Human Resources Laboratory and sponsored by the Air Force Office of Scientific Research. The goals of the program are to specify the basic parameters of learning ability, to develop techniques for the assessment of individuals' knowledge and skill levels, and to explore the feasibility of a model-based system of psychological assessment.

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 of Universal Energy Systems, Wesley Regian, Dan Woltz, and William Alley of the Air Force Human Resources Laboratory, and Stephanie Rude, Bruce Britton, and Shawn Glynn of the University of Georgia, for their valuable comments on the taxonomy and this report.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

What is the relationship between intelligence and learning ability? This question engaged contributors to the original *Learning and Individual Differences*, and we believe (and hope to show how) the sophistication of the answer to this question highlights, perhaps as clearly as to any other question, exactly how far our theories have come over the last 20 years.

Until recently, and certainly in evidence throughout that previous volume, the typical response to such a question might very well have been "there is no relationship between intelligence and the ability to learn" or "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 (e.g., 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 issue, primarily because of problems with the measures of learning ability he employed: His learning tasks may have been too simple (Campione, Brown, & Bryant, 1985; Humphreys, 1979) 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 (e.g., Snow, Kyllonen, & Marshalek, 1984).

A general conclusion may be drawn 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 point of being able to 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[1].

[1] 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 is logically prior to establishing the individual differences dimensions underlying learning. Proposing a learning skills taxonomy should assist in determining the dimensions of learning ability. We realize that our use of the terms *abilities* and *skills* may be somewhat idiosyncratic.

Indeed, there are many potential benefits to having a widely accepted taxonomy of learning skills. Consider Bloom's (1956) *Taxonomy of Educational Objectives*. 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 for evaluating student achievement consistent with those goals. Although the taxonomy has been criticized for vagueness (Ennis, 1986), it has served teachers well over the last 30 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 scientific and practical, in which a performance taxonomy in psychology would be beneficial. The main scientific benefit would be that results from different studies using differing 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 one 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 (e.g., a particular algebra curriculum) or outside the classroom (e.g., 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 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 areas (e.g., Fredericksen & Pine, in press). It is likely that, due to 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 would be realized in the burgeoning domain of intelligent computerized tutoring systems (ITSs). A number of such systems have been developed (Yazdani, 1986), and the potential for generalizing and synthesizing results across the different systems is seen as increasingly critical (Soloway & 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 (e.g., Anderson, Boyle, & 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.

The intelligent tutoring system context is a natural beneficiary of 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 were simply not addressable in the past. Educational research has been 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 ITSs, 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. So, a test of the utility of any learning taxonomy is whether it could be used to actually assist in such an endeavor. Our goal for this chapter is to propose such a taxonomy. We begin by looking at what has been done thus far.

II. A TAXONOMY OF LEARNING TAXONOMIES

Various approaches to the development of learning taxonomies have been employed. One way of organizing these approaches, which we apply here, is by the categories of (a) designated/rational, based on a conditions-of-learning analysis; (b) empirical-correlational, based on an individual differences analysis; and (c) model-based, from formal computer simulations of learning processes.

Designated/Rational Taxonomies

Designated/rational taxonomies are by far the most common. Examples of this type are taxonomies proposed by Bloom (1956), Gagne (1965; 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 was updated 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 Gagne (1965; 1985). Jensen proposed a three-faceted taxonomy: a *Learning-type* facet incorporated Melton's seven categories; a *Procedures* facet indicated variables such as the pacing of the task, stage of learning, whether the task consisted of spaced or massed practice, and the like; and a *Content/Modality* 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.

Gagne's taxonomy (1965; 1985), on the other hand, has been widely taught and put to use in the area of instructional design (Gagne & Briggs, 1979). Gagne proposes five major categories of learned capabilities based on a rational analysis of common performance 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 the 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 Gagne included to round out the list.

These categories serve various purposes. They assist the investigator in defining and analyzing instructional objectives during task analysis, and later, 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 Gagne, the five capabilities differ as to the conditions most favorable for their learning. For example, with verbal information, order is not important but providing a meaningful context is; for motor skills, providing intensive practice on part skills is critical.

All of these taxonomic systems--Gagne's in particular--are beneficial, but it is important to acknowledge their limitations. One problem inherent in the rational 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 should 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.[2]

Empirical-Correlational Taxonomies

A second approach, 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 (e.g., Thurstone, 1938), and there have been some attempts to develop similar taxonomies of learning tasks (e.g., Allison, 1960; Malmi, Underwood, & Carroll, 1979; Stake, 1961; Underwood, Boruch, & Malmi, 1978).

The empirical-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

[2]It is historically interesting that it was at Melton's (1964) conference that Fitts (1964) 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.

on task Y, then a natural proposal is that a high proportion of the skills task X requires 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 testbeds 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 et al, 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.

The 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 learning, memory span, frequency judgment). It was expected that clear word-attribute factors would emerge, thus supporting certain theoretical notions regarding properties of memory, but 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-by-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 vs. paired-associates and serial learning); but even this apparently was not a robust task division. A followup study (Malmi et al.,

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 vs. 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; thus, their tasks represented 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 Gagne's taxonomy. In this regard, we consider a pair of studies by Allison (1960) and Stake (1961), who administered a diverse variety of learning tasks to large samples of Navy recruits and seventh-graders, 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 & Snow, 1977), but a reanalysis conducted by Snow et al. (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 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, & Snow, 1983; Tversky & Hutchins, 1986).

Snow et al. 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 vs. familiar learning task dimension, which Snow et al. (1984) interpreted as supporting the classical distinction between fluid and crystallized intelligence (Cattell, 1971), but which might also be seen as supporting an inductive vs. rote learning distinction. 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 vs. 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 Snow et al. (1984) reanalysis thus provides a number of ideas that could facilitate taxonomy development. In particular, it suggests task complexity and learning environment (inductive/novel vs. rote/familiar) 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 & Snow (1977) could be reanalyzed along the lines of the Snow et al. approach. Even with the well-designed studies Snow et al.

reanalyzed, there is considerable under-determination of process dimensions, due to the fact that not enough varieties of learning tasks were (or could have been) administered by Allison (1960) and Stake (1961). Thus, although the dimensions revealed in the Snow et al. reanalysis 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 empirical-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 to-be-analyzed battery in the first place? Factor-correlational structures or categories directly reflect the nature of the tasks included in the analysis--and only those tasks; 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 task relationships that might not have been anticipated at the outset. But it cannot be complete in any sense. One cannot simply be sure 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 schools of thought--behaviorism in the case of rational taxonomies, psychometrics in the case of the empirical-correlational taxonomies--that are historically prior to 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 (e.g., Anderson, 1983; Rosenbloom &

Newell, 1986; Rumelhart & 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 empirically based individual differences taxonomies in that they have not yet been completely validated, at least not as taxonomies of learning skills. However, we do see a correspondence between some of the dimensions that have emerged in the individual differences 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 operational 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* 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 Underwood et al.'s (1978) finding 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" or 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 on novel problems to the employment of specialized, task-specific skills tuned to the particular problem at hand. Anderson's third learning mechanism, *strengthening*, operates somewhat analogously to tne 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 empirically based individual difference dimension and a distinction implicit in the model-based taxonomy. Snow et al.'s novel learning tasks, presumed to tap fluid intelligence, may be likened to Anderson's novel learning situations, which presumably tap very general problem-solving skill. On the other side, Snow et al.'s 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 vs. model-based approaches, there is some convergence in the categories of learning skills. Although Anderson (1983; 1987) 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. Taxonomies of research in machine learning (Carbonell, Michalski, & Mitchell, 1983; Langley, 1986; Michalski, 1986; Self, 1986) have been proposed, 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 et al. and Michalski suggest are listed in Table 1. (We have added an additional category, Learning by Drill & Practice, to the list, because we use the list as the basis for one of our proposed taxonomy categories, and it is convenient to denote that here.) Note that while there may be some similarity between Carbonell et al. and Michalski's categories and those proposed by Melton, Gagne, and others, the basic difference is the fact 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 (1983) and Carbonell-Michalski's (1983; 1986) model-based attempts to propose varieties of learning represent an advance beyond either the rational or empirically based taxonomies and go a long way toward 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 this paper represents our initial attempt to integrate these ideas.

III. 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. Our goal for this section of the paper is 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, or operate a camera, or 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

Table 1. *Learning Strategies From a Taxonomy of Machine Learning Research*

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

Learning from Instruction: 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 "coarse-grained" problem-solving search.

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

Inductive Learning: Learning by drawing inductive inferences from facts and observations obtained from a teacher or an environment.

Learning by Analogy: Mapping information from a known object or process to a less know but similar one.

Learning from Examples: Inferring a general Concept Description from examples and (optionally) counterexamples of that concept.

Learning from Observation & Discovery: 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 Deductive Learning (Michalski, 1986) are from Carbonell et al. (1983). The definitions are taken from the glossary in Michalski, Carbonell, and Mitchell (1986). *Learning by Drill and Practice* was not a category included in these sources, but we included it in the taxonomy and thus, for economy, we describe it here.

as what kinds of individual differences to expect and its relation to other cognitive skills. We know considerably less about procedural learning skills. The taxonomy may pinpoint other learning skills on which research attention may productively be focused.

We have selected four dimensions, illustrated in Figure 1, as particularly important in classifying learning skills. The two dimensions shown in Figure 1a--knowledge type and instructional environment--are motivated primarily by our discussion of the Anderson and Carbonell-Michalski systems, respectively, although Gagne's ideas on learned capabilities served to broaden the range of categories included in knowledge type. The crossing of these two dimensions (Figure 1a) defines a space of general learning tasks.

The motivation for the other two dimensions, illustrated in Figures 1b and 1c--*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 1b illustrates a hypothetical *domain-space* as the crossing of the degree of quantitativeness and the importance of quality vs. 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 1a) may prove to be empirically distinct from parallel learning skills in other domains. We represent this idea in Figure 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 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 declarative-procedural distinction is fundamental. Further refinements are possible; declarative knowledge can be arrayed by complexity, from propositional knowledge to schemata (packets of related propositions). Similarly, procedural knowledge can be arrayed from simple

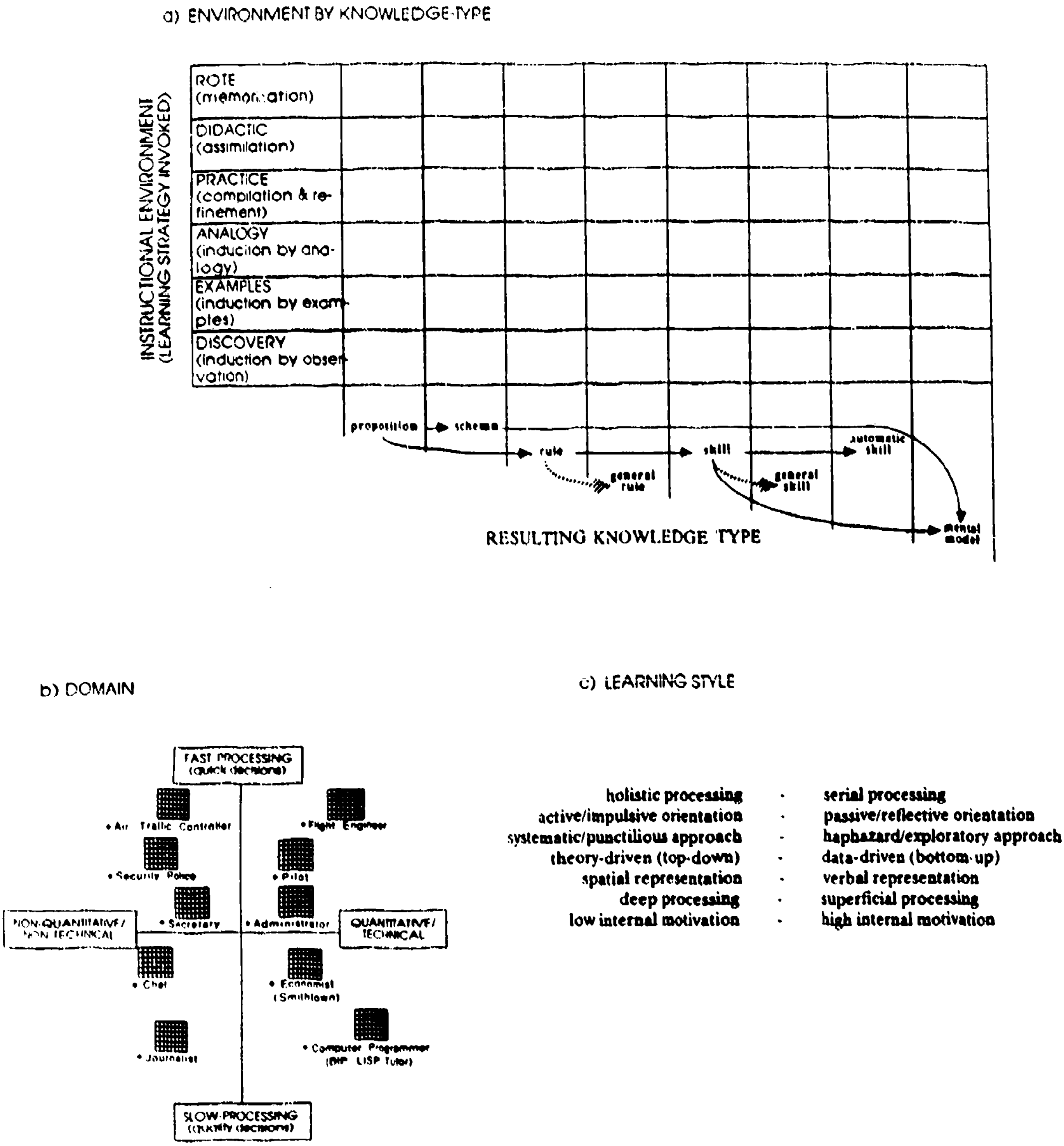

Figure 1. Learning skills taxonomy; a) Environment by knowledge-type matrix: cell entries would be various learning tasks; b) Environment by knowledge-type matrices plotted in a hypothetical two-dimensional domain-space: proximal matrices should show relatively greater transfer among parallel learning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

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, etc.), or in the case of schemata, with clustering and sorting tasks (e.g., Chi, Feltovich, & 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 pattern (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, & Vidulich, 1983) or through other methods of increasing processing demands (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; Spelke, Hirst, & Neisser, 1976). Possession of an appropriate mental model might require testing performance on a complex simulation of some target task. An illustrative (not exhaustive) list of tests for the various knowledge types is given in Table 2.

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

Table 2. *Sample Tests for the Various Knowledge Types (from the Domain of Logic Gate Circuits)*

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

diverse environment from the standpoint of student vs. teacher control and consequently in the kinds of inferences students are required to make. Even in the lecture environment, students may engage a variety of inferencing strategies. 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 1), we propose to characterize local instructional environments according to the amount of student control in the learning process. At one end, rote learning (e.g., memorizing the times 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.

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) suggested that metacognitive skills might be fairly generalized. But even here, there is not much evidence that metacognitive skill in mathematics (Schoenfeld, 1985) predicts metacognitive skill in writing (Hayes & 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 one considers the universe of occupational training courses rather than simply academic training. As is suggested in Figure 1b, it may be that the relative importance of speed vs. quality in decision-making may be a critical domain dimension. But again, the dimensions portrayed in Figure 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 (i.e., 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 primarily to reflect ability (e.g., Cronbach & Snow, 1977; Linn & Kyllonen, 1981). Impulsivity-reflectivity (Baron, Badgio, & 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 vs. serial processing, activity level, systematicity and exploratoriness, theory-driven vs. data-driven approaches, spatial vs. verbal representation of relations (Perrig & Kintsch, 1984), superficial vs. deep processing, and low vs. high internal motivation. Some dimensions may affect learning outcomes quantitatively: Active students may learn more. Others may affect outcomes qualitatively: Spatial vs. 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 vs. serialist processing styles, can illustrate this interaction. In this study, serialists, who focus on low-order relations and remember information in lists, were contrasted with holists, who focus on high-order relations and remember the overall organization among items to be learned. Pask and Scott showed that presenting a learning task (i.e., 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 1a set in the domain-space of Figure 1b). Each cell represents a task that teaches a particular subject matter (e.g., physics principles: Newton's second law), by a particular means (e.g., by analogy), resulting in a particular kind of knowledge outcome (e.g., 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 non-quantitative 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, extracted from a programming text, of teaching the concept of *electric field* (Glynn, Britton, Semrud-Clikeman, & Muth, 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 paraphrase recognition or recall.[3]

The electric field concept could be instructed 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

[3]Interestingly, test-question type has been shown to determine a learner's subsequent processing strategy (Fredericksen, 1984; Sagerman & Mayer, 1987).

providing a particular kind of environment guarantees neither the kind of learning experience that will result nor the kind of learning skill being tapped. Person characteristic by instructional treatment interactions exist (Cronbach & Snow, 1977, especially Chapter 11); thus, as we tried to illustrate in the example on holist vs. serialist processing, the style engaged at the time of learning and testing will partly determine what learning skill is being measured.

IV. APPLYING THE TAXONOMY: THREE CASE STUDIES

Our goal for this section of the paper is 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. We have proposed a taxonomy; it is now appropriate to 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, though not the focus of the paper, is to help us classify instructional programs. By our taxonomy, similar programs are ones that teach the same type of knowledge (propositions, skills, etc.), provide the same instructional environment (rote, discovery, etc.), teach the same domain material (computer programming, economics, etc.), and encourage the same kind (style) of learner interaction (reflectivity, holistic processing, etc.). 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 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 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 before, and so forth. After this, the program both updates 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.

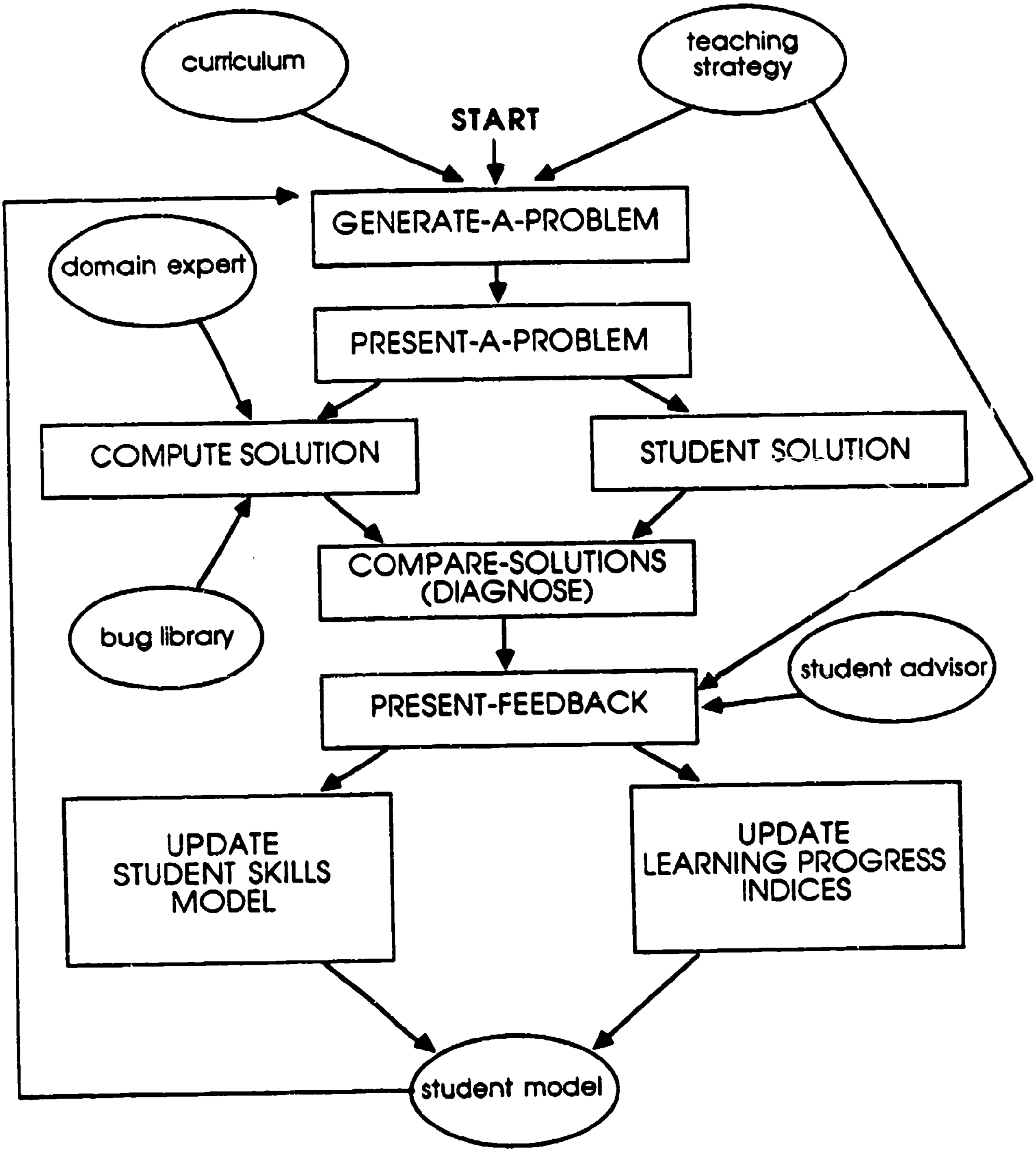

Figure 2. Components of a generic intelligent tutoring system. (Boxes represent decisions the program makes; ellipses represent knowledge bases the program consults.)

(1) BIP: Tutoring Basic Programming

General System Description

The Basic Instruction Program (BIP) was developed at Stanford University's Institute for Mathematical Studies in the Social Sciences and was one of the first operational intelligent tutoring systems (Barr, Beard, & Atkinson, 1976; Wescourt, Beard, Gould, & Barr, 1977).[4] BIP teaches students how to write programs in the language BASIC, by having the student solve problems of increasing difficulty. 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.

BIP's architecture is consistent with the generic tutor. BIP's *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 an example, printing a numeric literal (or constant) is considered conceptually analogous to, but also easier than, printing a string literal; both are considered easier than printing a numeric variable; and 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 12th 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.

[4]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.

BIP's student model is a list of the student's status with respect to each of 93 skills in the curriculum. There are five discrete status levels: UNSEEN (student has not yet seen a problem that required the skill), TROUBLE (student has seen but has not solved a problem that required the skill), MARGINAL (student has learned to a marginal degree), EASY (student has not yet seen but problem requires an easy skill to learn), and LEARNED (student has 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 which catches syntax errors, and a solution evaluator which determines whether the program is producing correct outputs. Finally, BIP also provides a number of aids to the student. The student may request help, a model solution in flowchart form, or a series of partial hints.

BIP selects problems by first identifying skills for which the student is ready (ones that do not have any unlearned prerequisites) but that need work, which means (in order of priority) (a) skills which the student has found difficult (i.e., from tasks not completed), (b) skills analogous to LEARNED skills, or (c) skills postrequisite to LEARNED skills. Skills so identified are called 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 (i.e., 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 29 subjects who had used BIP, and performed a number of analyses on the relationships among those data and BIP variables. Table 3 shows the list of learning indicators used by Snow et al. We have divided the list into three categories: learning progress 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 (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 to the activity variables than was the raw number of skills acquired. Snow et al. (1986) suggested this may have been due to the skills slope's 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 usage 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.

Table 3. *Learning Indicators from BIP, the Programming Tutor*

LEARNING PROGRESS 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

1. Planning: Proportion of time spent before coding
2. Implementing: Proportion of time spent writing code
3. Debugging: Proportion of time spent debugging code

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

Applying the Taxonomy

In evaluating the BIP tutor with respect to the taxonomy, we ask two questions: (a) What learning skills does BIP exercise (i.e., how can BIP be classified)? and (b) How comprehensive are the indicators used by Wescourt et al. (1977) and Snow et al. (1986) in measuring students' learning skills and their learning progress?

To address the first question, consider a distinction between what is tested and what is taught. BIP primarily tests for fairly specific skills, in that virtually all its tests are of the multiple operator selection variety (i.e., students write programs). The posttest also undoubtedly taps some propositional, schematic knowledge, but not extensively. Other knowledge outcomes could be tested, 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 & 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 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, indicated as the striped bar. Bar size represents the proportion of time spent either engaging the learning skill (solid) or having the skill 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 students engaged in the various activities, based on a review of Snow et al.'s (1986) data on the learning indicators, and on Wescourt et al.'s (1977) report 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 of the time spent exercising and testing various learning skills, separately for each student. Also note that only the knowledge type and instructional environment dimensions are indicated in Figure 3. Domain is indicated in Figure 1b (computer programming is highly quantitative/technical and quality of decisions is emphasized). Learning style is not directly assessed in BIP.

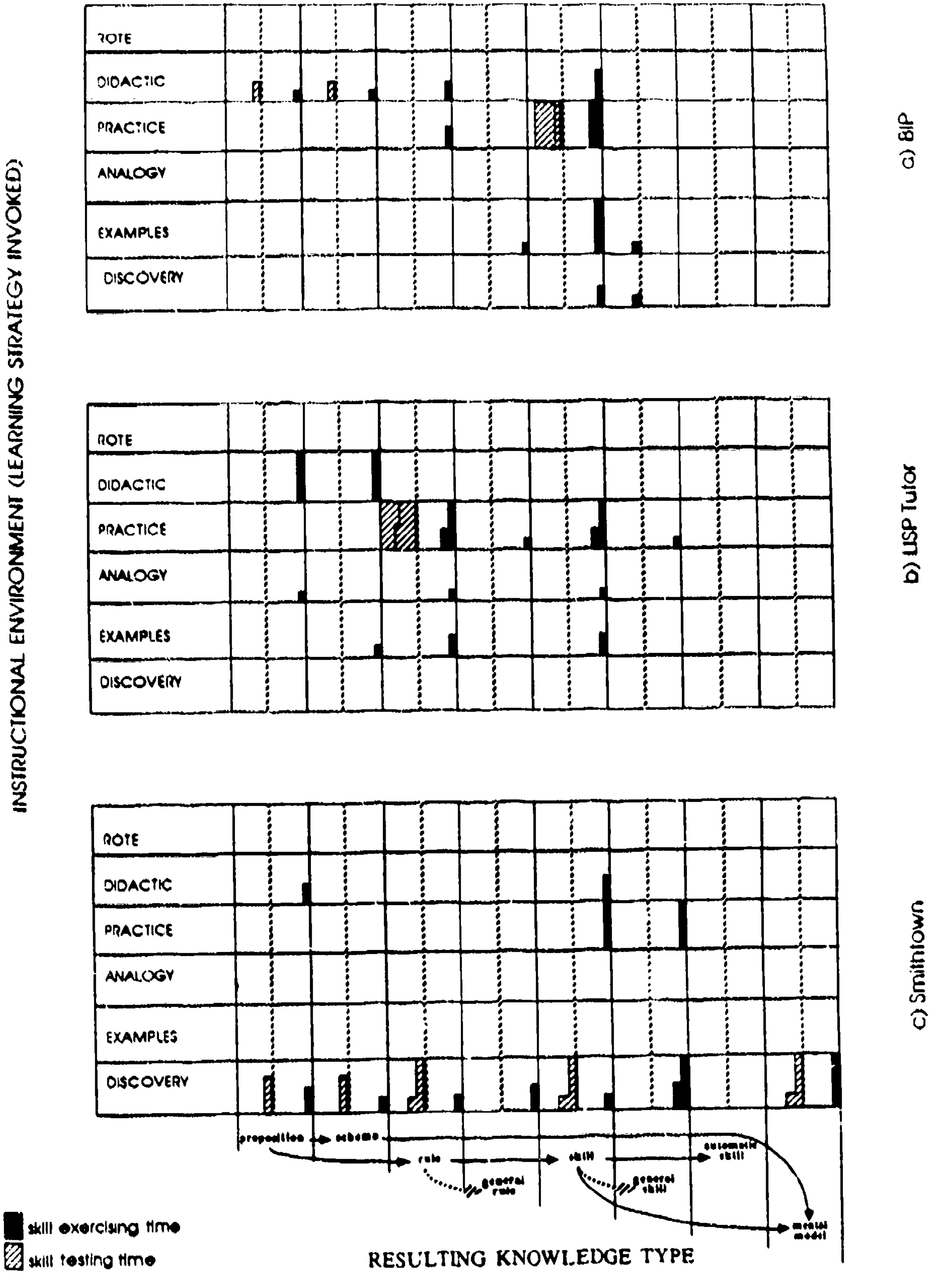

Figure 3. **Learning activities profiles for a) BIP, b) the LISP tutor, and c) Smithtown; solid bars represent the proportion of time the particular skill (defined by the environment x knowledge type cell task) is exercised by the tutor, relative to other skills; striped bars represent the proportion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

An approach to the second question, concerning indicator comprehensiveness, is suggested by Figure 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 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 test the propositional and schematic knowledge resulting from reading the text more directly by administering sentence verification tests, sorting tasks, and the like (see Table 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 a student's 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 appropriate-level problems. 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. But we think that the taxonomy can be used to provide an additional psychological basis for expressing those indicators.

(2) 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, et al., 1985); 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) called 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% 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 or 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 10 lesson-sessions (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 was performed on these data to determine whether production factors would emerge. For example, it could be that productions associated with one kind of learning (e.g., learning to trace functions, planning) would form a factor separate from some other kind of learning (e.g., learning to select functions, 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 a student's skill without directly evaluating that skill (i.e., 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 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 fact that the LISP tutor is capable 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 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 for 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 are 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 derives from the fact that the LISP tutor's multi-faceted coaching capability, which offers various kinds of tutorial remediation, greatly expands 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 the question of whether instruction using analogies results in greater subsequent ability to use the rule(s) so instructed, for example.

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.

(3) 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 & 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, defines 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, and then coaching students to become more effective problem solvers. The system keeps a detailed history list of all student actions, grouping them into (i.e., 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 non-actions) that constitute nonoptimal behaviors (e.g., "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 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 (e.g., "As price increases, quantity demanded decreases"), the student is congratulated and told the name of the law just discovered (e.g., "Congratulations! You have just discovered what economists refer to as the *Law of Demand*").

Learning Indicators

Shute, Glaser, and Raghavan (in press) conducted an extensive evaluation of differences among students in what the students 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 shows a set of 29 learning indicators constructed for analyzing individuals' performance. Indicators are clustered into three general behavior categories: (a) *activity - exploratory level skills* (indicators relating to activity level and exploratory behaviors), (b) *data management level skills* (indicators for data recording, efficient tool usage, and use of evidence), and (c) *thinking and planning level skills* (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 for which ones discriminated successful from unsuccessful learners. Two subjects--one who performed

Table 4. *Learning Indicators from Smithtown, the Economics Tutor*

ACTIVITY/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

Table 4. *Learning Indicators from Smithtown (cont.)*

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)

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 (i.e., effective experimental behaviors). In particular, the better subject collected and organized his data from a more theory-driven perspective, which contrasted with the 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 within a given market, and did not move randomly among markets as did the 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 her 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 he 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 (i.e., learning 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 3c as learning propositions and skills by direct instruction. Figure 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 online indicators used by the system in diagnosis, and (b) the separate posttest that measures economics knowledge gained during the tutorial. For the purpose of filling out Figure 3c, we considered half the total testing to be online and the other half to be the posttest; the striped bars are marked as to the testing source. Figure 3c shows that as in the LISP tutor, the online 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 assess students' ability to run mental simulations of complex economics scenarios (see Shute & Glaser, in press, for a detailed discussion of the test).

Figure 3c suggests that perhaps the greatest mismatch between what learning skills were exercised and what 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 similar 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 economics principles was tested at posttest, but not while students were interacting with the tutor. It seems reasonable, both from a research standpoint and from 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 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, a key question addressed in the Shute et al. analysis had to do with style interrelationships (the "dimensionality of style" question) and the relationship between style and learning outcome (the validity question). 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.

V. 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. (1986) generated and analyzed a set of learning indicators, Anderson (in press) 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, has to do with the true reduced-space dimensionality of learning skills (see footnote 1). 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 for instructional programs learning indicators 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/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 program of basic research designed to explore the possibility of using contemporary cognitive theory as the basis for a new system of ability measurement (Kyllonen, 1986; Kyllonen & 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 & Christal, in press, for current reviews), but the prototypical study investigates the relationship among 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 importantly, the learning tasks we have employed have tended to be short-term laboratory tasks, and therefore may not be truly representative of real-world learning activities. 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 laboratory has recently been built at Lackland Air Force Base that accommodates 30 work stations capable of administering intelligent computerized instruction such as 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 relationships among 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 5 and 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.

VI. 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 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 Gagne's

Table 5. *Applications of the Taxonomy: Suggestions for Design*

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/mental models to be taught.

   c. Determine tests to be used for assessing particular knowledge outcomes (Table 2).

2. Determine environment for achieving knowledge outcomes:

   a. Consider the kind of learning strategy desirable to invoke (Table 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 3 and 4). This also allows more precise statements of the effectiveness of the insruction.

   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.

Table 6. *Applications of the Taxonomy: Suggestions for Analysis*

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 2).

   c. Determine the match between the tests used and the knowledge outcomes intended (as in Figure 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 1) during the exchange. Generate learning activities profiles for the entire instructional program (as in Figure 3).

   b. Organize records of student learning success with respect to the knowledge-outcome-by-instructional-environment (KO x IE) matrix. That is, devise a means for assigning each student a separate learning success score for each cell in the KO x 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 x 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 x IE profiles separately 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 x IE matrices (possibly for separate styles), and domain.

Table 6. *Applications of the Taxonomy (cont.)*

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 activities profile, Figure 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.

Table 7. *Applications of the Taxonomy: What It Can Be Used For*

INSTRUCTIONAL SYSTEM EVALUATORS

(Teachers and Administrators)

- Facilitates analysis of what kinds of learning skills are being exercised and tested in an instructional setting (see Figure 3)

INSTRUCTIONAL SYSTEM DESIGNERS

- Suggests a range of possible instructional environments for achieving particular knowledge outcomes (see Table 1/Figure 1)

- Specifies techniques (tests) for probing a wide range of knowledge and learning skill outcomes (see Table 2)

COGNITIVE RESEARCHERS

- Suggests predictions about transfer relations among learning experiences (see Figure 1/Table 6)

- Suggests indicators (dependent variables) of what and how well a student is learning (see Figure 3/Tables 2, 6) taxonomies do. The advantage to 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 routine instructional settings. A second use for the system concerns primarily the environment dimension. The specification of multiple instructional environments permits the assessment of a range of means for achieving particular knowledge outcomes. If an instructor's goal is to teach a mental model of some system, the instructor can simply instruct it, or use an analogy, or have the student discover the model through observation of the system, and so on. 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 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 is also important for evaluating educational systems. Consider a general problem in evaluating innovative educational programs (discussed by Nickerson, Perkins, & 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 depend on standard achievement outcome tests as a means for their validation. Though 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. Additionally, a wide range of basic research issues emerges. 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.

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, MA: Harvard University.

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, & R. Neches (Eds.), *Production system models of learning and development*. Cambridge, MA: MIT Press.

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

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

Baron, J., Badgio, P. C., & Gaskins, I. W. (1986). Cognitive style and its improvement: A normative approach. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Volume 3* (pp. 173-220). Hillsdale, NJ: Erlbaum.

Barr, A., Beard, M., & 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., & Bryant, N. R. (1985). Individual differences in learning and memory. In R. J. Sternberg (Ed.), *Human abilities: An information processing approach* (pp. 103-126). New York: Freeman.

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

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

Cronbach, L. J., & 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 & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and Practice* (pp. 9-26). New York: Freeman.

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* (pp. 244-285). New York: Academic Press.

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

Fredericksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*, 193-202.

Fredericksen, N., & Pine, J. (in press). Assessment of learning in science and mathematics. *Report on Indicators of Quality of Education in Science and Math*. Washington, DC: National Academy of Sciences/National Research Council.

Gagne, R. M. (1964). Problem solving. In A. W. Melton (Ed.), *Categories of Human Learning* (pp. 294-317). New York: Academic Press.

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

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

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

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

Glynn, S. M., Britton, B. K., Semrud-Clikeman, M., & Muth, K. D. (in press). Analogical reasoning and problem solving in science textbooks. In J. A. Glover, R. R. Ronning, & 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., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, NJ: 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* (pp. 117-135). Columbus, OH: 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. & Christal, R. E. (in press). Cognitive modeling of learning abilities: A status report of LAMP. In R. Dillon & J. W. Pellegrino (Eds.), *Testing: Theoretical and applied issues*. San Francisco: Freeman.

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

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

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

Marshalek, B., Lohman, D. F., & 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* (pp. 325-339). 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* (pp. 238-252). Columbus, OH: Merrill.

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

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.) *Machine learning: An artificial intelligence approach, Volume II*. Los Altos, CA: Morgan Kaufmann.

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

Nickerson, R. S., Perkins, D. N., & Smith, E. E. (1985). *The teaching of thinking*. Hillsdale, NJ: Erlbaum.

Pask, G., & Scott, B. C. E. (1972). Learning strategies and individual competence. *International Journal of Man-Machine Studies*, *4*, 217-253.

Perrig, W., & Kintsch, W. (1984). *Propositional and nonpropositional representations of text* (Tech. Rep. No. 131). Boulder, CO: Institute of Cognitive Science, University of Colorado.

Reiser, B. J., Anderson, J. R., & Farrell, R. G. (1985). Dynamic student modeling in an intelligent tutor for LISP programming. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Volume 1* (pp. 8-14).

Rosenbloom, P. S., & Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach, Volume II*. Los Altos, CA: Morgan Kaufmann. Michalski.

Rumelhart, D. E., & Norman, D. A. (1981). Analogical processes in learning. In J. R. Anderson (Ed.). *Cognitive skills and their acquisition* (pp. 335-359). Hillsdale, NJ: Erlbaum.

Sagerman, N, & Mayer, R. E. (1987). Forward transfer of different reading strategies evoked by adjunct questions in science text. *Journal of Educational Psychology*, *79*, 189-191.

Schneider, W., & 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., & 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., & Glaser, R. (in press). An intelligent tutoring system for exploring principles of economics. In R. E. Snow & D. Wiley (Eds.). *Straight thinking*. San Francisco: Jossey Bass.

Shute, V. J., Glaser, R., & Raghavan, K. (in press). Discovery and inference in an exploratory laboratory. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences*. San Francisco: Freeman.

Snow, R. E., Kyllonen, P. C., & 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, NJ: Erlbaum.

Snow, R. E., Wescourt K., & Collins, J. (1986). *Individual differences in aptitude and learning from interactive computer-based instruction* (Tech. Rep. No. 10). Stanford, CA: Stanford University School of Education.

Soloway, E., & Littman, D. (1986). Evaluating ITSs: The cognitive science perspective. *Proceedings of the Research Planning Forum for Intelligent Tutoring Systems*. San Antonio, TX: Air Force Human Resources Laboratory.

Spelke, E., Hirst, W., & 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. & 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., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General*, *107*, 393-419.

Wescourt, K. T., Beard, M., Gould, L., & Barr, A. (October, 1977). *Knowledge-based CAI: CINS for individualized curriculum sequencing* (Tech. Rep. No. 290). Stanford, CA: Institute for Mathematical Studies in the Social Sciences, Stanford University.

Wickens, C. D., Sandry, D., & 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 & 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.