5. What knowledge engineering or task analysis methods are required to support your approach?

We are indebted to a lot of people for making the conference and resulting volume a reality. First, we would like to acknowledge the contributions of Mike Young and Tom Killion for starting the Cognitive Skill Acquisition Workshop several years ago. We also extend heartfelt thanks to Mark Miller and Systems Explorations, Incorporated. Mark handled all of the logistics, large and small, that were involved with the organization of the conference and the volume. We applaud and appreciate the time and toil of all the speakers/contributors. Without them, of course, there would have been no conference or book. We truly appreciate the help of our astute friends who took time to review and critique final drafts of these chapters: Scott Chaiken, Ray Christal, Lisa Gawlick, Pat Kyllonen, and Bill Tirre.

J. W. Regian
V. J. Shute
San Antonio, Texas

1 Automated Instruction as an Approach to Individualization

J. Wesley Regian
Valerie J. Shute
Armstrong Laboratory, Brooks Air Force Base, Texas

We are interested in *automated instruction,* instruction that is delivered on any microprocessor-based system. The term as we use it may include, but is not limited to, computer-assisted instruction, computer-based training, intelligent tutoring, simulator-based training, interactive videodisk-based training, computerized part-task training, and embedded training. We believe that it is possible for automated instructional systems to be more effective than they currently are. Specifically, we believe that by using artificial intelligence programming techniques, it is possible for automated instructional systems to emulate the desirable properties of human tutors in one-on-one instruction.

Gamble and Page (1980; see also O'Neil & Baker, 1991) speculated that effective human tutors:

1. cause the heuristics of the student to converge to those of the tutor;
2. choose appropriate examples and problems for the student;
3. can work arbitrary examples chosen by the student;
4. are able to adjust to different student backgrounds;
5. are able to measure the student's progress; and
6. can review previously learned material with the student as the need arises.

Automated instructional systems have been built with various subsets of these capabilities (see Wenger, 1987). Moreover, we believe it is possible, in principle, for automated instructional systems to surpass the instructional effectiveness of human tutors due to certain inherent properties of automated systems. Automated systems are relentless in their persistence, being unable to "burn out" or become unmotivated to help the student succeed. Automated instructional systems have the capability to graphically and behaviorally simulate desired learning and transfer contexts.[1] Finally, such systems can simulate psychomotor aspects of the transfer context (such as with simulators and virtual world environments). The goal of this book is to outline a set of principled approaches to tapping the potential of automated instructional systems to individualize instruction.

THE PROMISE OF INDIVIDUALIZED INSTRUCTION

Three overlapping streams of research provide the historical opportunity for this book: (a) research into individualized instructional approaches often called mastery learning approaches, (b) research into interactions between subject variables and instructional treatments called aptitude-treatment interactions, and (c) research into advanced computer-based instructional systems called intelligent tutoring systems (ITS). The common thread through these research streams is the belief that individually tailored instruction is superior to group-oriented instruction.

The idea that teaching is best accomplished by tailoring instruction to individual students is both ancient and ubiquitous among instructional theorists. Corno and Snow (1985) found the idea detailed in the 4th century B.C. Chinese Xue Ji, in the ancient Hebrew Haggadah of Passover, and in the 1st century Roman De Institutione Oratoria. Today the basic idea still forms the core of several important streams of research on instruction. The promise of individualized instruction is the basis of research on mastery learning (e.g., Bloom, 1956; Carroll, 1963; Cohen, Kulik, & Kulik, 1982), aptitude-treatment interactions (e.g., Corno & Snow, 1985; Cronbach & Snow, 1977; Shute, chapter 2, this volume, in press), and intelligent tutoring systems (e.g., Burton & Brown, 1982; Lewis, McArthur, Stasz, & Zmuidzinas, 1990; Woolf, 1987). The idea also has strong empirical support. A great deal of data indicates that carefully individualized instruction is superior to conventional group instruction (Bloom, 1984; Woolf, 1987). A consistent finding is that when using traditional stand-up instruction, other things being equal, smaller class sizes produce superior learning outcomes. The most common interpretation of this result is that smaller classes enable instructors to be more aware of, and responsive to, the instructional needs of individual students.

There is not full agreement on the best way to be responsive to the needs of individual students. The research streams introduced here represent relatively distinct ways to think about individualized instruction. In one approach, individuals are thought to differ primarily on learning rate. For example, enthusiasts

[1]Examples include jet engines, nuclear reactors, orbital dynamics, principles of economics, laws of physics, and steam turbines.

of programmed instruction (Skinner, 1957) believed that learning rate was the only individual difference worthy of attention. They believed that over time, however, learning-rate differences produce differences in readiness to learn because of failure to have learned previous material in the allotted time. Mastery learning enthusiasts seem to have been heavily influenced by this school of thought (see Carroll, 1963).

In a second approach to individualized instruction, learners are thought to differ on various dimensions, or aptitudes. According to this way of thinking, an aptitude is any characteristic of the individual that is supportive of future achievement in some situation. Thus, aptitudes may be learned or innate. Aptitude-treatment interaction researchers (see Cronbach & Snow, 1977) look for interactions between these aptitudes and instructional approaches, or treatments.

In a third approach to individualized instruction, individuals are characterized in terms of information-processing models of task performance. Instructional interventions are planned after comparing a model of trainee performance (the student model) to a model of expert performance (the expert model). Alternatively, no effort may be expended in developing a student model. In this case, instructional interventions are planned after comparing student performance to the expert model (see Anderson, 1990).

The three approaches (rate, aptitude, model-based) make different assumptions about how to design and implement individualized instruction. The research agenda for rate theorists is to make teachers and students aware of the student's progress toward mastery and to provide each student with ample time to learn. The research agenda for aptitude theorists is to identify the critical aptitudes along with the instructional approaches that are most suitable for levels of these aptitudes. And finally, model-based theorists work to develop models of expert performance and then identify the best approach to moving the student's performance toward the expert model. There are hybrid approaches. Shute (chapter 2, this volume), for example, describes an intelligent tutoring system using a combination of model-based and aptitude approaches. In the following sections we review each of the three approaches in greater detail.

MASTERY LEARNING

The Approach. Several innovative models of classroom instruction originating in the 1970s sought to allow teachers in group instructional settings to approximate individualized instruction from a slightly elaborated rate perspective. In general, these programs were based on a combination of clear instructional objectives and periodic diagnostic evaluations, allowing teachers to be aware of and responsive to students' knowledge levels throughout the learning process (Stallings & Stipek, 1986). The most influential of these programs were Learning for Mastery (LFM; Bloom, 1968) and the Personalized System of

Instruction (PSI) or the Keller plan (Keller, 1968). These, and related programs, came to be known collectively as mastery learning. The basic premise of mastery learning programs is as follows: Rather than holding instructional time constant and allowing achievement to vary, it is better to hold achievement constant and allow instructional time to vary. That is, given enough time and appropriate intervention, most or all students can achieve mastery of the instructional objectives. Further, if a student is not allowed to achieve mastery of current instructional objectives, then he or she is certainly doomed to failure on later instructional objectives that are hierarchically built upon current instructional objectives.

Contributors to the mastery learning research stream typically measure the effectiveness of an instructional intervention by comparing the mean of the treatment (mastery learning) group to the mean and standard deviation of the control (conventional instruction) group on some final measure of performance. For example, a 1 sigma effect would mean that the average student in the treatment group was 1 SD above the average student in the control group on the final performance measure. This final performance measure may be a standardized achievement test but is usually an experimenter-generated local measure of performance on specific objectives.[2]

Retrospective. In a series of studies, Bloom (1984) reported two important findings about mastery learning and individualized instruction. First, under mastery learning conditions, students performed an average of 1 *SD* above traditionally instructed students, or at the 84th percentile (see Fig. 1.1). Second, under individual tutorial conditions, students performed an average of 2 *SDs* above traditionally instructed students, or at the 98th percentile (see Fig. 1.1). Bloom challenged researchers to develop mastery-based methods of achieving the 2-sigma effect in group teaching situations. These data, and the 2-sigma challenge, are the subject of some controversy (see especially Slavin, 1987; Kulik, Kulik, & Bangert-Drowns, 1990). The controversy centers on the question of whether such a large effect size can ever be consistently obtained in traditional instructional settings using standard achievement tests as criterion measures.

Slavin (1987) conducted a meta-analysis of 17 controlled evaluations of mastery learning programs. In 13 of those studies that looked at experimenter-made measures of summative performance rather than standard achievement tests, the effect size ranged from −0.11 to 0.90 sigma with a mean of 0.34 sigma. Slavin argued that Bloom's (1984) 2-sigma challenge is unrealistic and that his 1-sigma claim is based on studies that are too brief and too small. Although Slavin is seen by mastery learning enthusiasts as a detractor (see, e.g., the emotional reply to Slavin, 1987, by Anderson & Burns, 1987), he actually believes a 0.33-sigma effect to be realizable in traditional instructional settings using standard achievement tests as criterion measures. Further, he pointed out that an effect of that magnitude would wipe outs the average achievement gap between lower- and middle-class children in just 3 years. He called for continued research to achieve the potential of mastery approaches in practical application.

[2]There is little evidence of any mastery learning effect on standard achievement tests. This is a bone of contention to mastery learning detractors. It is not, however, damaging to the goal of developing instruction that is targeted to clear and specific objectives. There is good evidence that mastery learning can work in such cases.

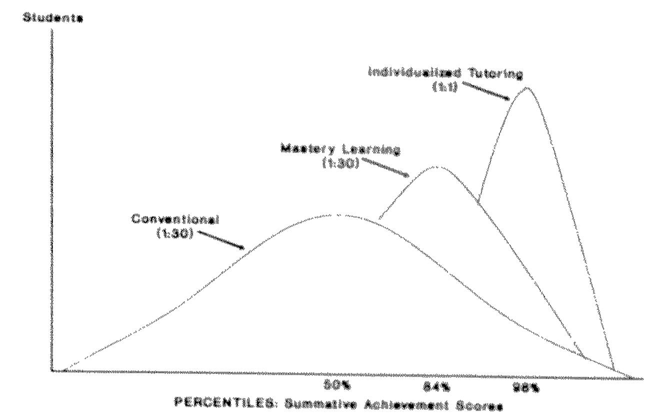

FIG. 1.1. Distributions for different learning conditions (adapted from Bloom, 1984).

Kulik et al. (1990) conducted a meta-analysis of 108 controlled evaluations of mastery learning programs. They looked at 36 evaluations of Bloom's LFM approach and 72 evaluations of Keller's PSI in college, high school, and upper elementary school settings. In these studies the effect size ranged from 0.22 to 1.58 sigma with an average of 0.52 sigma. Thus, the average student in the mastery learning condition performed at the 70th percentile on the summative evaluation, as compared to the 50th percentile for students in the control condition.

Finally, Shute (in press-b) reported the effect sizes from two evaluation studies conducted with intelligent tutoring systems (ITS): one teaching avionics troubleshooting—Sherlock (Lesgold, Lajoie, Bunzo, & Eggan, 1990), and one teaching scientific inquiry skills—Smithtown (Shute & Glaser, 1991). A 1-sigma effect size was computed for both tutors when two groups of learners were compared: learning the curriculum with and without the respective computer programs. These ITS evaluations are thus in the same league as the Bloom mastery learning data (i.e., 84th percentile).

Lessons Learned. A review of mastery learning research shows that it is possible to use diagnostic tests and ongoing remediation during instruction to produce large enhancements in instructional effectiveness over traditional instruction. Apparently, however, the effect is limited to situations where the goal is to teach clear and specific objectives. Cronbach and Snow (1977) took strong issue with the claim that individual differences in achievement can be eliminated by varying instructional time. There is no convincing evidence that it is possible to eliminate or significantly reduce individual differences on standard achievement tests. Individual differences can, however, apparently be reduced significantly if achievement refers to specific performance objectives rather than general achievement.

Mastery enthusiasts say little about how to effectively diagnose student learning and say nothing about optimal remediation. This probably accounts for the large variability in the effect size. Overall, mastery approaches are poorly implemented in that teachers are not provided with necessary training, resources, or assistance (Slavin, 1987). For example, they are often provided with no training on how to create corrective instruction resulting from diagnostic information, and often the corrective instruction is just given too late. In some cases, remediation is provided as late as 4 weeks after diagnosis (Slavin, 1987). The mastery learning enthusiasts seem to have fallen prey to a pitfall pointed out by Cronbach and Snow, "What one cannot do is generalize about instructional techniques in the abstract" (Cronbach & Snow, 1977, p. 214). Thus, we see a great need to develop guidelines and principles to drive diagnosis and remediation.

Most mastery learning research has been done in the classroom under uncontrolled or partially controlled conditions (Block, 1974). For example, of several hundred studies under consideration for a meta-analysis, Slavin (1987) only deemed 17 to be rigorous enough for inclusion in his final set. Also, due to the bias for publishing positive results, there is no way to access information about mastery learning programs that failed to produce a treatment effect.

For several reasons it is difficult to implement mastery learning on a mass scale. Levin (1974) pointed out that in group settings, individual diagnostic information is costly, and even when it is obtained, knowledge of how to select individualized treatments is speculative.

APTITUDE-TREATMENT INTERACTION

The Approach. The goal of ATI research is to relate the selective effectiveness of various instructional treatments to measurable characteristics of individuals. The relationship can take the form of capitalization, remediation, or compensation (Cronbach & Snow, 1977). The treatment can capitalize on assets, preferences, or tendencies of the individual; compensate for weaknesses; or remediate shortcomings. Aptitude refers to any measurable characteristic of the individual that is propaedeutic to achievement in a given situation (Corno & Snow, 1986). Thus, aptitudes may include knowledge, skills, abilities, personality characteristics, attitudes, and so on.

Retrospective. Opinions about ATI are polarized. Bracht (1969) reviewed 90 studies and concluded that ATI were found only as often in these studies as would be expected by chance. Glass (1970), commenting on Bracht's review of ATI studies, said: " 'There is no evidence for an interaction of curriculum treatments and personological variables.' I don't know of another statement that has been confirmed so many times by so many people" (p. 210).

Cronbach and Snow (1977), in the final chapter of their book on ATI, also commented on the Bracht review. In addition, they reviewed much of the same literature that Bracht did and more, although using a finer level of analysis:

> Aptitude × Treatment interactions exist. To assert the opposite is to assert that whichever educational procedures is best for Johnny is best for everyone else in Johnny's school. Even the most commonplace adaptation of instruction, such as choosing different books for more and less capable readers of a given age, rests on an assumption of ATI that it seems foolish to challenge (p. 492).

ATI involving general ability are far more common in the literature than ATI involving more specific abilities. A common finding is that when comparing fully elaborated treatments to student-directed treatments, highs (on general ability) profit from student-directed treatments, whereas lows are handicapped. Cronbach and Snow believe that highs "profit from the opportunity to process the information in their own way" (p. 500). This is consistent with the finding that curriculum preorganizers are useful for lows and sometimes detrimental to highs. There is some evidence that lows can be helped through the use of clarifying demonstrations or devices without damaging the performance of highs.

There are occasional interactions involving more specialized abilities such as spatial and mathematical abilities, and with other aptitudes such as prior learning, memory, and personality styles. Across studies looking for these effects, however, conflicting results are found more often than not. It is probable that there is fertile ground here for cultivation, but principles are not forthcoming as yet.

Lessons Learned. ATI have been very difficult to find because of insufficient sample sizes, poor methodology, various uncontrolled conditions, and unanticipated interactions that occur across settings. Cronbach and Snow (1977) argued for several methodological rules of thumb in designing ATI research. Most studies of ATI are brief and artificial, pointing to a "need to collect data from instructional procedures that realistically progress through a body of material" (p. 509).

Cronbach and Snow believe that time should be held constant, allowing achievement to vary, rather than carrying each subject to criterion and allowing instructional time to vary. They specifically referred to both the mastery learning and the programmed instruction streams of research as often violating this rule

Their argument rested strongly on the assumption that the end-user of instructional research is the traditional educational system. The recommendation is less critical for industrial or military training because the trainee becomes employable as soon as he or she reaches criterion on task performance.

ATI studies with random assignment to one of two treatments should use about 100 subjects per treatment. This rule of thumb can be relaxed somewhat for sufficiently powerful designs involving extreme groups or matched cases. Most investigators in the ATI tradition before 1977 used 40 or fewer subjects per treatment, and may have lacked the power to pick up even moderate effects.

Cronbach and Snow warned against oversimplification in research about ATI, arguing strongly for the measurement of multiple aptitudes, treatments, and outcomes. They also warned of the complexities of research in intact classes and naturalistic educational research.

Finally, Cronbach and Snow argued that the choice of treatment conditions should be principled, based on a detailed taxonomy of instructional situations, and on process analysis or other theoretical approaches to performance.

INTELLIGENT TUTORING SYSTEMS

The Approach. The introduction of artificial intelligence technology to the field of computer-aided instruction (CAI) has prompted research and development efforts in an area known as intelligent computer-aided instruction (ICAI). We may conceive of computer-based training (CBT) systems as lying along a continuum that runs from CAI to ICAI. There are important differences between CAI systems and ICAI systems.

As we have discussed, a great deal of data indicates that under certain circumstances, diagnostically tailored instruction can be superior to untailored instruction. Thus, an important way in which CBT systems differ is in the degree to which their behavior is modified by an inferred "model of the student's current understanding of the subject matter" (VanLehn, 1986). The CBT system that is less intelligent by this definition may be conceived of as CAI. The system that is more intelligent may be conceived of as ICAI. Often, ICAI systems are referred to as intelligent tutoring systems (ITS; Sleeman & Brown, 1982). This term is particularly appropriate, as it brings to mind one-on-one tutoring.

With respect to individualization, it is important to note that virtually all traditional CAI systems are individualized in the sense that they are self-paced, and many are further individualized by virtue of branching routines that allow different students to receive different instruction. CAI systems with branching routines are, in fact, more individualized than those without branching routines. Thus, they are more intelligent by the current definition (although in a weak sense). In branched CAI the instructional developer must explicitly encode the actions generated by all possible branches, and there is a finite number of

possible paths through these branches. As we move further away from the CAI to the ICAI end of the continuum, we begin to see a very different and more powerful approach to individualization. This more powerful approach was touched on by Wenger (1987) when he referred to explicit encoding of knowledge rather than encoding of decisions. An ITS (which term probably should be reserved for systems that are very far toward the ICA end of the continuum) uses a diverse set of knowledge bases and inference routines to "compose instructional interactions dynamically, making decisions by reference to the knowledge with which they have been provided" (Wenger, 1987, p. 5). The "intelligence" in these systems resides in cognitive diagnosis—the ability to analyze learners' solution histories dynamically, using principles rather than preprogrammed responses to decide what to do next and how to adapt instruction to different learners.

Retrospective. Since the 1980s, many ITSs have been built incorporating various approaches to diagnostic student modeling and remediation. Far fewer systems have been formally evaluated, but of the subset of evaluated systems, some have approached the kind of instructional power produced by individualized human-taught instruction (e.g., Lesgold et al., 1990; Reiser, Anderson, & Farrell, 1985; Shute & Glaser, 1990).

Shute (in press-b) reviewed findings from four studies conducted with ITSs, some of the select few that have undergone empirical evaluation: (a) the LISP tutor, which teaches programming in LISP (Anderson, Farrell, & Sauers, 1984); (b) Smithtown, which teaches scientific inquiry skills in the context of microeconomics (Shute & Glaser, 1991); (c) Sherlock, which teaches avionics troubleshooting (Lesgold et al., 1992); and (d) the Pascal Tutor, which teaches programming in Pascal (Bonar, Cunningham, Beatty, & Weil, 1988; Shute, 1991a). The results of the evaluations were impressive. Learning efficiency (rate) with ITSs was accelerated in comparison to control conditions. Overall, students acquired the subject matter faster from various ITSs than from more traditional environments. For example, subjects working with the LISP tutor learned the knowledge and skills in one third to two thirds the time it took a control group to learn the same material (Anderson, Boyle, & Reiser, 1985). Subjects working with Smithtown learned the same material in half the time it took a classroom-instructed group (Shute & Glaser, 1990). Subjects working with Sherlock learned in 20 hours skills that were comparable to those possessed by technicians having almost 4 years experience (Nichols, Pokorny, Jones, Gott, & Alley, in preparation). And finally, subjects learning from the Pascal ITS acquired, in one third the time, equivalent knowledge and skills as learned through traditional instruction (Shute, in 1991b).

With regard to learning outcome, ITSs again performed well in comparison to control conditions. The LISP tutor group attained the same (or in one study, 43% better) criterion scores as a control group not using the tutor. Results from the

Smithtown analysis showed that subjects learned the same material as a classroom group, despite the fact that the tutor focused on the instruction of scientific inquiry skills, not the subject matter. The outcome data from subjects using Sherlock showed increases in scores comparable to an advanced group of subjects and significantly better than a control group. In all cases, individuals learned faster, and performed at least as well, with the ITSs as subjects learning from traditional environments. For a more thorough treatment of these evaluation studies, see Shute (1991b).

Lessons Learned. A review of ITS research suggests that it is possible to use artificial intelligence to develop computer-based instructional systems that automatically generate and deliver tailored instruction. This automatically tailored instruction can (at least sometimes) produce large enhancements in instructional efficiency or effectiveness over nontailored instruction.

Although it is accurate to say that most of the evaluation studies published to date have shown positive effects, this is misleading. In studies of instructional interventions, there is a selection bias for publication of effective interventions. Also, controlled evaluations of ITSs are rare (Baker, 1990; Littman & Soloway, 1988), even though there are many published accounts of ITS design and development (see Wenger, 1987). A review of these accounts shows that ITSs are often designed haphazardly, the range of domains for which they have been built is somewhat narrow, and implementation of system components is often guided by "intuition" rather than theory (e.g., Koedinger & Anderson, 1990; Norman, 1989). If the current generation of ITSs were subjected to controlled evaluation, the results would probably be quite variable.

THE FUTURE OF AUTOMATED INSTRUCTION

Mastery learning researchers have tried to iteratively move learners toward mastery of task performance. ATI researchers have tried to match up instructional treatments with measurable characteristics of individuals. ITS researchers have tried to move learners toward a well-specified expert model of performance. There are success stories in each of these research streams. However, general and systematic principles of individualized instruction have not emerged. Studies in these areas are plagued by noisy data, methodological flaws, small samples, and various unpleasant constraints arising from the realities of educational environments. Donchin (1989) described some of the problems as follows:

> As my colleagues and I examined the literature on training and practice we became increasingly, and painfully, conscious of the fact that it is very difficult to integrate the studies we were reviewing. The theoretical acumen and the ingenuity of previous investigators was beyond reproach. A vast number of papers had been published within such domains as "learning theory," "training," "motor behavior" and similar areas. However, it was quite evident that the diversity of paradigms and theoretical approaches within which the phenomena were studied, and the models tested, made it difficult to compare results across studies. The many contradictions which are frequent in any body of literature were difficult to resolve because much of the conflict could be attributed to the different settings, and paradigms, in which the phenomena were studied (pp. 4–5).

Many have argued that instructional research would benefit from a more systematic approach to pedagogy. We believe that progress can be enhanced by specifying instructional approaches clearly and in sufficient detail to allow others to apply, evaluate, and compare the approaches. The ensuing chapters in this volume outline a variety of approaches to automated instruction.

REFERENCES

Anderson, J. R. (1990). *Analysis of student performance with the LISP tutor. Diagnostic monitoring of skill and knowledge acquisition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Anderson, J. R., Farrell, R., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science, 8,* 87–129.

Anderson, L. W., & Burns, R. B. (1987). Values, evidence, and mastery learning. *Review of Educational Research, 57,* 215–223.

Baker, E. L. (1990). Technology assessment: Policy and methodological issues. In H. L. Burns, J. Parlett, & C. Luckhardt (Eds.), *Intelligent tutoring systems: Evolutions in design* (pp. 243–263). Hillsdale, NJ: Lawrence Erlbaum Associates.

Block, J. H. (1974). Mastery learning in the classroom: An overview of recent research. In J. H. Block (Ed.), *Schools, society, and mastery learning* (pp. 28–69). New York: Holt, Rinehart & Winston.

Bloom, B. S. (1956). *The taxonomy of educational objectives: Handbook 1: The Cognitive Domain.* New York: McKay.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2).

Bloom, B. S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Bonar, J., Cunningham, R., Beatty, P., & Weil, W. (1988). *Bridge: Intelligent tutoring system with intermediate representations* (Tech. Rep.) Pittsburgh, PA: Learning Research & Development Center, University of Pittsburgh.

Bracht, G. H. (1969). *The relationship of treatment tasks, personological variables and dependent variables to aptitude-treatment interactions.* Unpublished doctoral dissertation, University of Colorado, Denver.

Burton, R., & Brown, J. S. (1982). An investigation of computer coaching for informal learning activities. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 79–98). New York: Academic Press.

Carroll, J. (1963). A model of school learning. *Teachers College Record, 64,* 723–733.

Cohen, P. A., Kulik, J., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*(2), 237–248.

Corno, L., & Snow, R. E. (1985). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 605–629). New York: Macmillian

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

Donchin, E. (1989). The learning strategies project, *Acta Psychologica, 71*, 1–15.

Gamble, A., & Page, C. V. (1980). The use of artificial intelligence techniques in computer-assisted instruction: An overview. *International Journal of Man-Machine Studies, 12*, 259–282.

Glass, G. (1970). Discussion. In M. C. Wittrock & D. C. Wiley (Eds.), *The evaluation of instruction* (pp. 210–211). New York: Holt, Rinehart, & Winston.

Keller, F. S. (1968). "Goodbye, teacher . . ." *Journal of Applied Behavioral Analysis, 1*, 78–89.

Koedinger, K. R., & Anderson, J. R. (1990). *Theoretical and empirical motivations for the design of ANGLE: A new geometry learning environment*. Working notes: AAAI spring symposium series. Stanford University, Stanford, CA.

Kulik, C. L., Kulik, J., & Bangert-Drowns, R. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*, 265–299.

Lesgold, A., Lajoie, S. P., Bunzo, M., & Eggan, G. (1992). A coached practice environment for an electronics troubleshooting job. In J. Larkin, R. Chabey, & C. Cheftic (Eds.), *Computer assisted instruction and intelligent tutoring systems: Establishing communication and collaboration* (pp. 201–238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Levin, H. M. (1974). The economic implications of mastery learning. In J. H. Block (Ed.), *Schools, society, and mastery learning* (pp. 73–88). New York: Holt, Rinehart & Winston.

Lewis, M. W., McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). *Discovery-based tutoring in mathematics*. Working notes: AAAI spring symposium series. Stanford University, Stanford, CA.

Littman, D., & Soloway, E. (1988). Evaluating ITSs: The cognitive science perspective. In M. C. Polson & J. J. Richardson (Eds.), *Foundations of intelligent tutoring systems* (pp. 209–242). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nichols, P., Pokorny, R., Jones, G., Gott, S. P., & Alley, W. E. (in preparation). *Evaluation of an avionics troubleshooting tutoring system* (Tech. Rep.) Armstrong Laboratory, Human Resources Directorate, Brooks AFB, TX.

Norman, D. A. (1989). *The psychology of everyday things*. New York: Basic Books.

O'Neil, H. F., & Baker, E. L. (1991). Issues in intelligent computer-assisted instruction: Evaluation and measurement. In T. B. Gutkin & S. L. Wise (Eds.), *The computer and the decision-making process* (pp. 199–224). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shute, V. J. (1991a). Who is likely to acquire programming skills? *Journal of Educational Computing Research, 7*(1), 1–24.

Shute, V. J. (in press). A comparison of learning environments: All that glitters . . . In S. P. Lajoie & S. J. Derery (Eds.), *Computers as cognitive tools*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shute, V. J. (1991b). *Meta-evaluation of four intelligent tutoring systems: Promises and products* (AL-TP-1991-0040). Air Force Systems Command, Brooks Air Force Base, Texas.

Shute, V. J., & Glaser, R. (1990). Large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments, 1*, 51–77.

Shute, V. J., & Glaser, R. (1991). An intelligent tutoring system for exploring principles of economics. In R. E. Snow & D. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research, 57*, 175–213.

Sleeman, D. H., & Brown, J. S. (Eds.). (1982). *Intelligent tutoring systems*. London: Academic Press.

Stallings, J. A., & Stipek, D. (1985). Research on early childhood and elementary school teaching programs. In M. Wittrock (Ed.), *Research on teaching* (pp. 727–753). New York: Macmillan.

VanLehn, K. (1986). Student modeling in intelligent teaching systems. *Proceedings of the Research Planning Forum for Intelligent Tutoring Systems*. San Antonio, TX: Air Force Human Resources Laboratory.

Wenger, E. (1987). *Artificial intelligence and tutoring systems*. Los Altos, CA: Morgan Kaufman.