COGNIMETRICS
innovative assessment solutions

A new work-environment:

New Aptitudes Require New Measures.

Martin J. Ippel

CogniMetrics, Inc.
San Antonio, TX

Jac. N. Zaal

GITP International
Nijmegen, NL

Paper presented at the
46th Annual Conference
of the
International Military Testing Association (IMTA)
26-28 October 2004
Brussels, Belgium.

Introduction

Information Technology has gradually, but radically, reshaped the environment in which we work, communicate, and do business. These changes in our work environments put new demands on workers. Problem solving, information processing skills, and the ability to interact with intelligent tools are gradually replacing crystallized intelligence, clerical speed and accuracy.

The IT Aptitude Battery tests measure an aspect of general cognitive ability that has become increasingly important since the advent of Information Technology. That is, the ability to interact efficiently with systems that have a dynamics of their own. These systems have become an indispensable part of the fabric of post-industrial society and vary in complexity from MS Office software tools to complex information systems monitoring industrial processes. The cognitive skill required for using such devices is called procedural knowledge, or procedural skill.

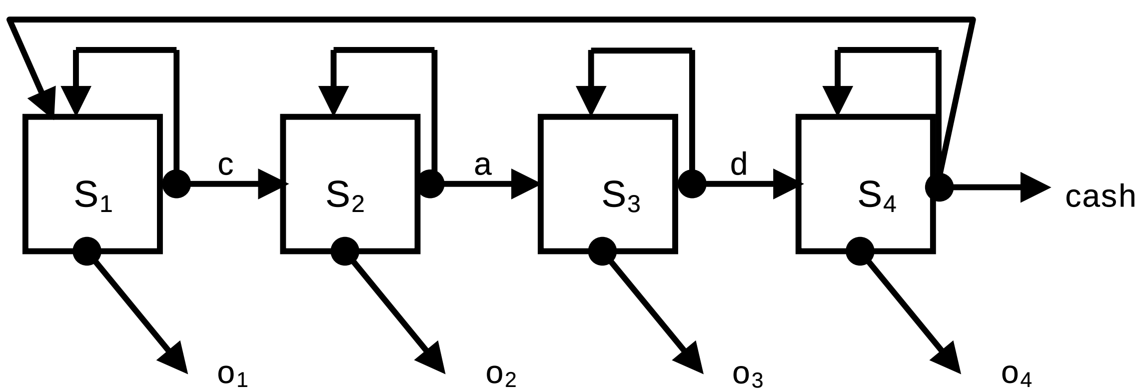


Figure 1. An ATM cash-out procedure.

States:	S_1, …, S_4, where
	S_1: identity unknown & amount requested unknown & account unknown & balance unknown;
	S_2: amount requested unknown & account unknown & balance unknown;
	S_3: account unknown & balance unknown;
	S_4: balance unknown;
Input Signals:	c (insert card), a (amount requested), d (determine account);
Output Signals:	o_1 (instruction), o_2 (question), o_3 (question), o_4 (evaluation information);

Procedural skills

Procedural knowledge is the knowledge that a person utilizes to interact with a physical, or a conceptual, system to achieve a particular goal. As an example of a simple procedure to interact with a physical system, imagine someone using an ATM to cash money. To achieve this goal this person must perform certain actions in response to output signals generated by the ATM. Figure 1 gives a schematic representation of the procedure to cash-out money from an ATM. To be able to do this successfully this person somehow should have acquired a representation of the function that governs the state transitions of this machine (Ippel and Beem, 1997). The ATM procedure is simple because each state transfers to the next upon a specific input action. The complexity of procedures can increase significantly when the same set of input actions have different effects depending on the state in which the system is. Examples of such more complex forms of procedural knowledge were described for the domain of computer programming (Ippel and Meulemans, 1998), and for remotely controlling a UAV payload camera (Ippel and Watson, 1998).

In general, a procedural skill comprises knowledge of the appropriate input signals (user actions), knowledge of how these signals affect the system's state transitions and knowledge of the output signals that indicate the system's current state so that the appropriate signal can be fed into the system when needed to achieve a particular goal.

Traditional tests of intelligence have a limited potential in tapping procedural knowledge. Within the traditional test paradigm it is very difficult, if not impossible, to capture the intricacies of procedural skills. Procedural skills typically consist of sequences of actions in response to state changes in a physical (or conceptual) system. To measure just the final result of these action sequences misrepresents the nature of procedural skills. As a result, traditional tests of general cognitive ability, or general intelligence, have limited capacity to assess the natural variation in aptitude for acquiring procedural knowledge (Ippel & Hurwitz, 1998).

Mechanisms Of Procedural Skill Learning

The present approach to measurement of procedural knowledge acquisition is based on recent results in the cognitive sciences with respect to development of procedural knowledge (e.g., Newell, 1990; Van Lehn, 1990). In Artificial Intelligence research as well as in cognition psychology thinking is modeled as searching through a state space - the search space. In psychology this search space is referred to as a problem space (Newell and Simon, 1972; Newell, 1990; Van Lehn, 1990). This problem space consists of a finite set of states, including an initial state and a goal state, and a finite set of (admissible) operators to transform one state into another.

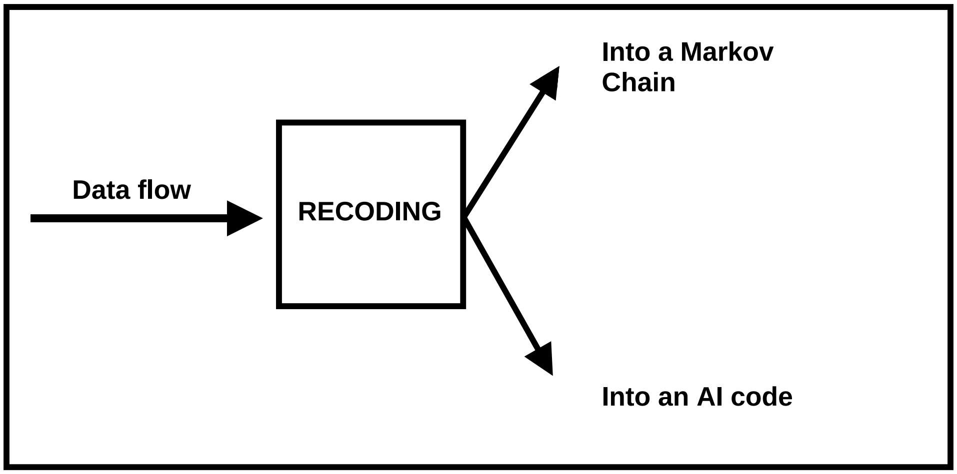
The basic mechanisms of procedural skill learning are (1) the repeated application of an elementary search control cycle (i.e., observe the system's current state, provide some action input, and observe the effect). The goal of the search is to find a proper sequence of actions that will transform the initial state of the system into the goal state, and (2) compilation, a process in which the search control knowledge increasingly becomes deterministic. The final result of compilation is that individual no longer searches for a sequence of operators, but executes a pre-existing program (i.e., a procedure) to achieve the goal.

The operation of this elementary search control cycle can vary from a completely random choice of new states to a completely deterministic choice of every following state, depending upon the (search control) knowledge of individual of the task environment. Procedural learning is thus modeled as increasing determinism in the individual's search control knowledge. When little knowledge is available, the learner has to rely on very general problem solving techniques (weak search methods). Although inefficient compared to the task-specific methods of someone familiar with the task, these weak search methods often provide the only basis for intelligent action (Ippel and Meulemans, 1990). Some examples are: generate-and-test, hill-climbing, means-ends analysis (Rich, 1983). While these methods have shown up first in AI investigations of problem solving, they seem to provide for a natural description of human problem solving behavior as well. There is evidence that even young children use such problem solving methods (e.g., Byrnes & Spitz, 1979; Borys, Spitz, & Dorans, 1982; Klahr, 1985).

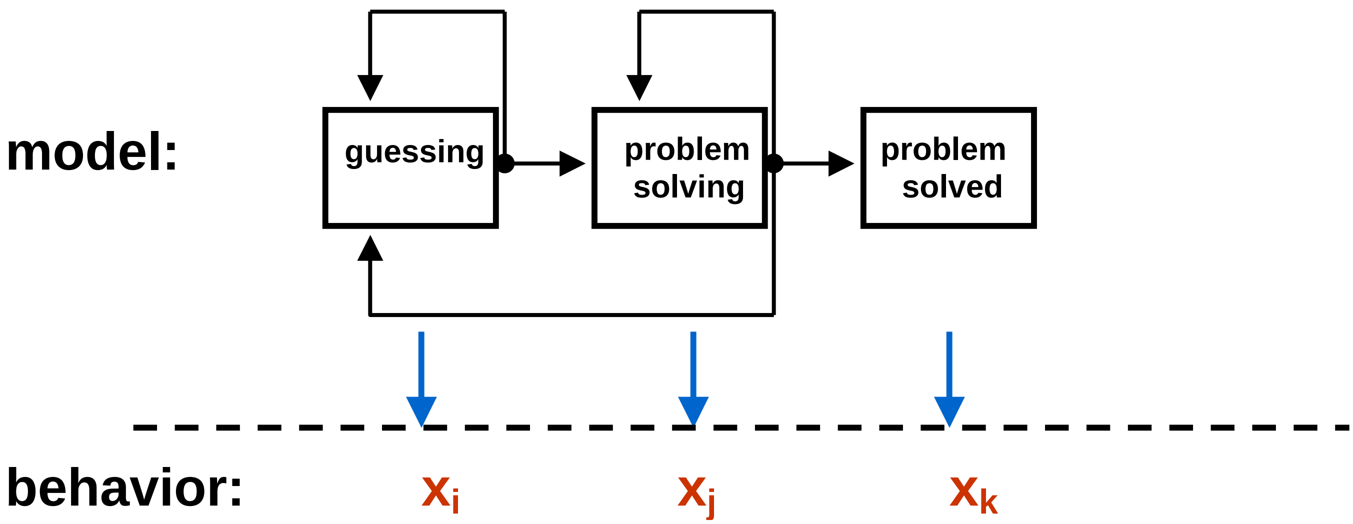
The IT Aptitude Battery consists of two tests, each designed in such a way that only one of these weak search techniques can be applied. The Hidden Target Test measures an individual's ability to solve problems and build an algorithm using a hill-climbing technique and the Battery Test measures the ability to use a means-ends approach to build an algorithm to solve a particular class of problems.

Characteristics of ITAB task environments

The ITAB test tasks were designed in analogy to an essential aspect of the job of a computer programmer / systems analyst, which is, to define an algorithm (or a set of algorithms) that can solve a well-defined class of problems. Each of the ITAB tests consists of a class of problems for which the examinee has to develop an algorithm that solves these problems efficiently. The tests measure how examinees incorporate feedback from the system into their follow-up actions and how quickly this leads to a build-up of an efficient algorithm.



A specific feature of the ITAB tests that the data stream is recoded into two separate codes. The first is a so-called Markov Chain code (analysis of action sequences) and the second is a so-called Artificial Intelligence code (analysis of "intelligence" of each step) (see Figure 2).

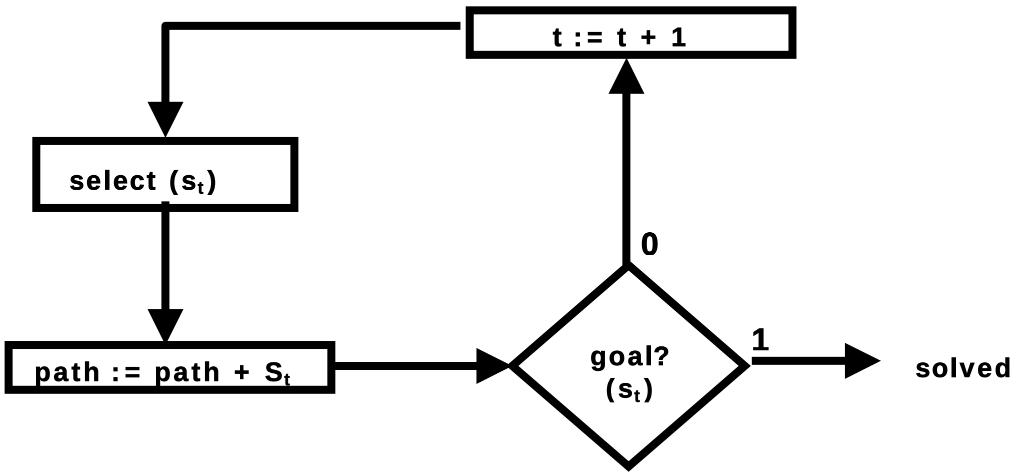


A Markov model is used to analyze the sequential patterns. A Markov model is a probabilistic model, which describes a process of state changes in a system. Figure 3 shows the Markov model used to analyze the sequential aspect of the Hidden Target Test data. The model is characterized by three states. The first state (G) is called GUESSING. The subject either has not received any feedback on an action and has to make a guess (first guess), or does not seem to incorporate the feedback into his action. The second state (P) is called PROBLEM SOLVING. The subject chooses his or her next action based on the previous feedback. The last state (S) is called PROBLEM SOLVED. Once the Markov process has arrived in this state, the process terminates. S is called an absorbing state; the other states are transient states.

ENT: General Performance Scores

The most important outcome of the Markov model analysis is the expected value of the number of steps to reach a solution for a particular individual (ENT-score). In Cognitive Psychology this parameter is often referred to as the *length of the solution path* (Newell & Simon, 1972; Newell, 1990) and is generally held to be indicative for the difficulty the individual is experiencing in solving a problem. Each ITAB test has an ENT-score, viz., ENT_H (for the Hidden Target Test) and ENT_B (for the Battery Test).

In order to get a better understanding of the nature of these scores, we will have a closer look at the mechanisms that generate those scores. For the sake of brevity, in this paper we will focus on the mechanism that generate the ENT_H score.



The Hidden Target Test generates a task environment in which the individual has to locate a hidden target in a search field. Upon each guess or test the system gives feedback in terms of distance of the guess or test to the hidden target. Figure 4.a shows the most elementary form of a so-called "generate-and-test" function (G&T method) that can produce a solution path. This G&T function generates a state randomly, and passes this state on to an evaluation function that subsequently tests whether it is a goal state (i.e., D (S_t) = 0). In this most elementary form no intelligence is involved. It is clear that this process may generate very long solutions paths (i.e., high ENT scores). Note the counter in Figure 4.a (i.e., t := t + 1) representing a counting / cataloguing process performed by a (minimal) memory function to prevent the system from selecting the same state every time.

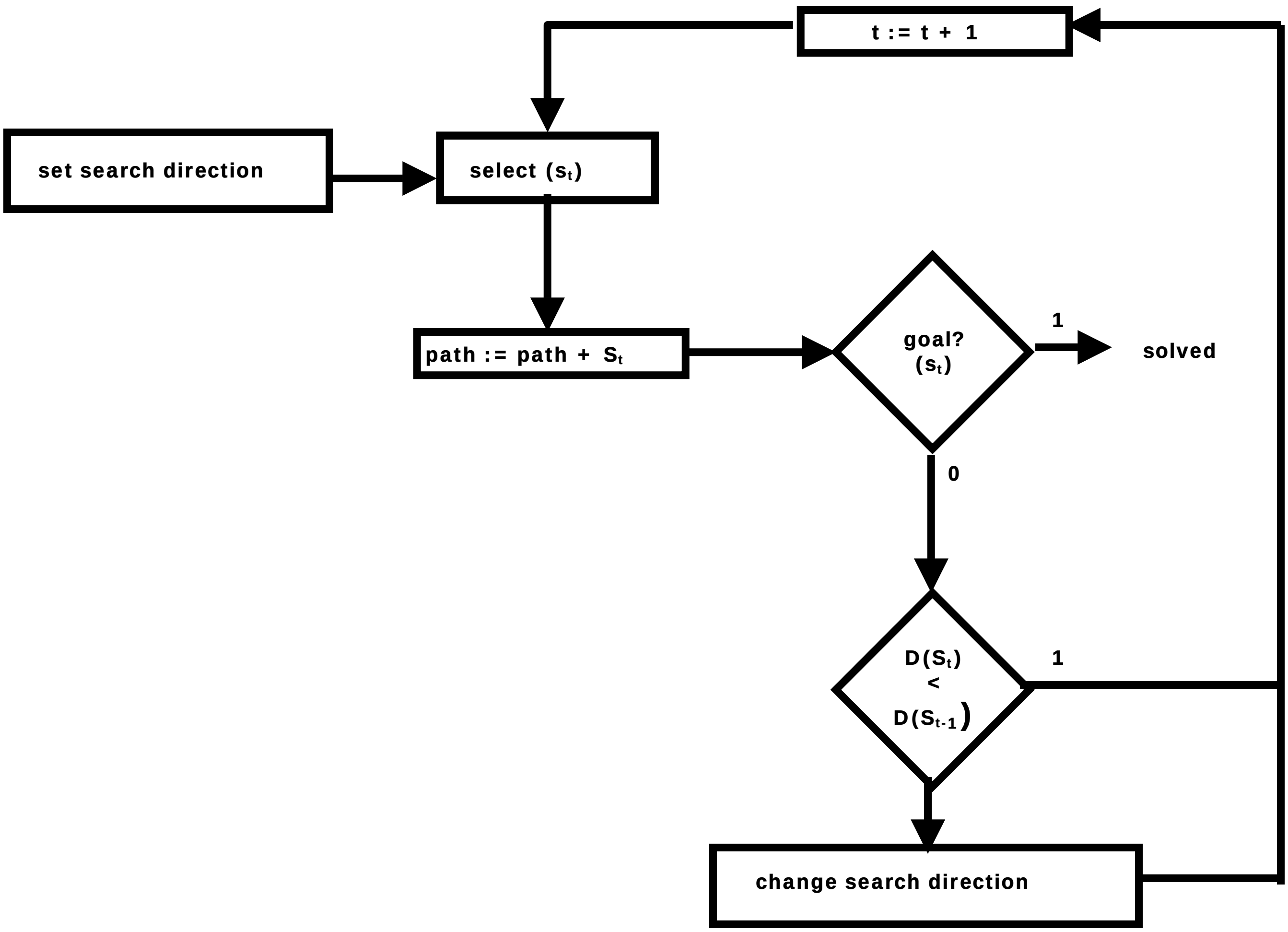


Figure 4.b. A hill-climbing method testing locations in the Hidden Target task environment.

Where does intelligence come in? In this process mechanism intelligence is usually modeled by augmenting the evaluation function in such a way that it provides an ordering principle for at least a subset of the states in the problem space. This method is known as hill-climbing (see Figure 4.b). Note that in an advanced form this process mechanism also requires an augmented memory function. In this advanced form his augmented memory should not only remember past states, but also be able to uphold simultaneously a representation of a (sub) set of states considered in one evaluation act.

In the Hidden Target Test the operation of this augmented evaluation function is measured in two independent ways. First, the Markov model analyzes the sequence of steps on one particular aspect: is step s_t closer to the goal state then s_{t-1}? (see second test box in Figure 4.b: $D(S_{t+1}) < D(S_t)$, where D = distance). If so, it is concluded that the individual is thinking. If not, it is concluded that he or she is guessing. Based on this analysis the Markov model estimates the expected value of the length of a solution path. Other more diagnostic outcomes of the analysis are estimates of the mean length of holding times for problem solving and for guessing.

Second, we were able to quantify the working of the heuristic function by applying Information Theory (Ash, 1965; Attneave, 1959; Krippendorff, 1986; Shannon & Weaver, 1949)). We distinguish two aspects in information processing: (1) How well does the individual use the information available in the feedback provided by the task environment? (2) How does this affect the amount of information processing per step?

- We measure the intelligence of each step in the problem solving process by comparing the information value of the next step chosen with the step that would generate the maximum value. The information value is measured in binary units per step (bits/step). The difference between the actual step and the best possible step defines an information loss. A maximally efficient problem solving process has a zero information loss. This variable is referred to as UHI in Table 1.
- Once the individual produced a step in the problem space, we can compare the level of uncertainty (H) before the step, that is, H (t-1) with the uncertainty after the step (H (t). A reduction in uncertainty level is a measure of information gained by that step. This variable referred to as IET in Table 1.

Table 1. PM correlations between HTT scores and WPT (N = 161).

	ENT-H	GT	PST	IET	UHI	RT
ENT-H						
GT	0.335					
PST	-0.480	-0.368				
IET	-0.604	-0.769	0.433			
UHI	0.380	0.918	-0.464	-0.785		
RT	0.480	-0.230	-0.035	-0.094	-0.160	
WPT	-0.357	-0.247	0.379	0.360	-0.265	-0.235

Since the IET and UHI variables are purely mathematical constructs, and true by definition, they can be used to validate the interpretation of the ENT-score and of the holding time variables derived from the Markov model.

Table 1 shows the results of a study on a sample that was representative for the adult US population. The results show a strong correlation between ENT and IET (r = 0.604, df = 160), suggesting that a shorter solution path goes with higher levels of information extraction per step. Information loss correlates at a more moderate level with the ENT-score (r = 0.380, df = 160). An important finding supporting the consistency of the conceptual framework of the test is the finding that higher levels of loss of feedback information strongly decrease the amount of information extracted per step (r(IET, UHI) = -0.785, df = 160). The correlations of the information measures with the estimated cycle lengths of the Markov model states are consistent with expectations.

Table 2. P.M. Correlations with External Variables.

	ENT-H	ENT-B
WPT	-0.357	-0.201
WMC	-0.281	-0.356
Conscientiousness	0.123	0.092
Extraversion	-0.061	-0.112
Neuroticism	0.065	0.048
Openness	-0.042	-0.041
Agreeableness	-0.036	-0.016
age	0.095	0.288
gender	-0.189	-0.042
yrs of education	0.002	0.123
N =	161	144

Correlations with external variables

Table 2 shows correlations of both ENT-scores with a classical measure of general intelligence (Wonderlic Personnel Test) and a composite score of two experimental Working Memory Capacity tests of the US Air Force (WM-S1 and WM-V1).

Table 3. Reliability Coefficients for several HTT scores.

	G-C[1]	SH[1]	n
ENT-H[2]	0.969	0.979	190
RT	0.932	0.973	188
IET	0.880	0.900	188
UHI	0.927	0.988	188

[1] G-C = Guttman-Cronbach's alpha; SH = Split-Half Coeff.
[2] Estimate based on analysis of number of steps per trial.

Reliability of the ENT-scores

Table 3 reports Guttman-Cronbach alpha values and split-half reliabilities for several Hidden Target Tests discussed so far. Table 4 reports the BT counterparts of those indices.

In general, the values presented in Table 4 were somewhat lower than those presented in Table 3. The reason is probably an error in the test administration: the time per trial was erroneously set at 1 minute per trial instead of two minutes per trial.

Table 4. Reliability Coefficients for several BT scores.

	G-C[1]	SH[1]	n
ENT-B[2]	0.792	0.792	128
RT	0.873	0.884	128
Hrem	0.835	0.848	128
Errors	0.868	0.881	128

[1] G-C = Guttman-Cronbach's alpha; SH = Split-Half Coeff.
[2] Estimate based on analysis of number of steps per trial.

Discussion

Developments in Information Technology have a twofold effect on human performance testing. First, the work-environment is changing and creates the need for new skills and aptitudes to acquire those skills. For example, cognitive skills required in dealing with interactive devices (e.g., database systems, programming environments, semi-automated systems). Second, Information technology has a unique potential to advance the field of human performance assessment to domains of cognitive functioning that previously were not accessible. For example, traditional tests do not capture the intricacies of procedural skills adequately. Even when technical knowledge is the measurement objective, it is being treated as declarative knowledge (e.g., the ASVAB, the primary test battery for selection and classification of the US Military). Information technology makes it possible to design tests capture the dynamics of behavior, and enables a high-density sampling of cognitive process indices.

CogniMetrics' IT Aptitude Battery embodies a set of innovations based on the possibilities of Information technology. Essential elements of the Hidden Target Test are: (1) the test provides an interactive environment, and (2) actions of the examinee are not scored as singleton answers to distinct problems, but are analyzed as sequence patterns. Complete interactivity is achieved by creating an internal representation of the task-environment. Artificial Intelligence technology is used to compute the "intelligence" of each step taken by the examinee.

This paper focuses on only one of the ITAB tests: the Hidden Target Test for reason of brevity and because this test is more intuitively accessible than the Battery Test, the other ITAB test. The conclusions hold for both tests.

References:

Ash, Robert, B. (1965). *Information Theory*. New York: Dover Publications.

Attneave, F. (1959). *Applications of Information Theory to Psychology.* New York: Holt.

Borys, S., Spitz, H.H., & Dorans, B.A. (1982). Tower of Hanoi Performance of Retarded Young Adults and Non-Retarded Children as a Function of Solution Length and Goal State. *British Journal of Experimental Child Psychology, 33,* 87 – 110.

Byrnes, M.A., & Spitz, H.H. (1979). Developmental Progression of the Tower of Hanoi Problem. *Bulletin of the Psychonomic Society, 14,* 379 – 381.

Ippel, M.J., & Beem, A.L. (1997). *Mental Models As Finite-State Machines: Examples and Computational Methods*. Air Force Research Laboratory Technical Report, AL-HR-TR-1997-0043.

Ippel, M.J. & Hurwitz, J.B. (1998). A New Generation of Tests. *Individual Differences in Performance*, Human Factors and Ergonomics Society, 4 –6.

Ippel, M.J. & Meulemans C.J.M. (1990). A Computer Simulation of the Acquisition of a Computational Skill in a Discovery-Oriented Microworld. In J.M. Pieters, K. Breuer, & Simons, P.R.J. (Eds.). *Learning Environments: Contributions from Dutch and German research (pp. 177-192*). Heidelberg: Springer-Verlag.

Ippel, M.J., & Meulemans, C.J.M. (1998). A Cybernetic Approach To The Study of the Learnability of the LOGO Turtle World. *Joural of Educational Computing Research, Vol. 19,* 2, 191 – 221.

Ippel, M.J., & Watson, S.J. (October, 1998). *The UAV Camera Directory Task As A Finite-State Machine.* Proceedings of the Human Factors and Ergonomics Society 42st Annual Meeting: Chicago (Il).

Klahr, D. (1985). Solving Problems with Ambiguous Subgoal Ordering: Preschoolers Performance*. Child Development, 56,* 940 – 952.

Krippendorff, K. (1986). *Information Theory. Structural Models for Qualitative Data.* London: Sage Publications.

Newell, A. (1990). *Universal Theories of Cognition.* Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H.A. (1972*). Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Rich, E. (1983). *Artificial Intelligence.* London: McGraw-Hill.

VanLehn, K. (1990). *Mind Bugs. The Origins of Procedural Misconceptions.* Cambridge, MA: MIT Press.

Shannon, C.E., & Weaver, W. (1949). The Mathematical Theory of Communication. Urbana: University of Illinois Press.