ARI Research Note 87-65

AD-A190 645

ISSUES IN
PSYCHOLOGICAL RESEARCH AND APPLICATION
IN TRANSFER OF TRAINING

Virginia E. Holt (editor)
American Psychological Association

for

Contracting Officer's Representative
Judith Orasanu

BASIC RESEARCH LABORATORY
Michael Kaplan, Director

DTIC
SELECTED
JAN 2 1 1988
E

ari

U. S. Army

Research Institute for the Behavioral and Social Sciences

December 1987

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

WM. DARRYL HENDERSON
COL, IN
Commanding

EDGAR M. JOHNSON
Technical Director

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 87-65	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Issues in Psychological Research and Application in Transfer of Training		5. TYPE OF REPORT & PERIOD COVERED Final Report September 86 - February 87
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Victoria E. Holt (editor)		8. CONTRACT OR GRANT NUMBER(s) TCN 86-647
9. PERFORMING ORGANIZATION NAME AND ADDRESS American Psychological Association 1200 Seventeenth Street, NW, Washington, D.C. 20036		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE December 1987
		13. NUMBER OF PAGES 79
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office) - -		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE n/a

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

- -

18. SUPPLEMENTARY NOTES

Judith Orasanu, contracting officer's representative.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Transfer of Training	Job Aids
Performance Measure	Training
Jobs Performance	
Skill Training	

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

This research note is a collection of papers and summary recommendations resulting from a two-day workshop focused on training transfer. Supported by the Army Research Institute, the workshop featured presentations by academic, non-academic, and military laboratory scientists on psychological research and applications related to transfer of training. Among the specific topics dealt with are: the development of cognitive simulation models, skills

(OVER)

DD $_{1 JAN 73}^{FORM}$ 1473 EDITION OF 1 NOV 65 IS OBSOLETE

ARI Research Note 87-65 (continued)

20. Abstract (continued)

development methods, the need for intelligent job aids, cognitive task anal-
vsis, and methods for measuring job performance. Recommendations for further
research and applications are provided.

Accession For		
NTIS GRA&I		☒
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution/		
Availability Codes		
Dist	Avail and/or Special	
A-1		
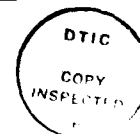
DTIC
COPY
INSPECTED

ISSUES IN PSYCHOLOGICAL RESEARCH AND APPLICATION IN TRANSFER OF TRAINING

Summaries from the

DEPARTMENT OF DEFENSE RESEARCH ROUNDTABLE

February 27-28, 1986

TABLE OF CONTENTS

PREFACE

This document is the result of a collaborative effort of the Army
Research Institute for the Behavioral and Social Sciences (ARI), the
American Psychological Association (APA), and the Federation of Behavioral,
Psychological and Cognitive Sciences (FBPCS).

The APA, with the cooperation of FBPCS, organized a workshop on
transfer of training issues as the first of a series of Department of
Defense Research Roundtable meetings. ARI provided travel funds for the
presenters and support, through its Scientific Services Program, for the
preparation of this summary document.

The idea for the Research Roundtable series was generated by the APA
Committee on Research Support (CORS), a committee reporting to APA's major
scientific policymaking body, the Board of Scientific Affairs. CORS
initiated the project to provide a forum for the discussion of behavioral
research issues related to the defense mission. Transfer of training was
chosen as the first topic for consideration.

CORS set the framework for the Roundtable project, and appointed
members William Howell and Gary M. Olson to provide oversight.
Subsequently, a planning committee developed specific guidelines for the
workshop and selected presenters. Planning Committee members were Gary M.
Olson, Irv Goldstein, David Kieras, and William Montague. Throughout the
process, Captain Paul Chatelier, Assistant for Training and Personnel
Technology, Office of the Deputy Assistant Secretary for Research and
Technology, and Milton Katz, ARI Director of Basic Research, provided
guidance and suggestions.

The individuals selected as presenters represent a spectrum of
interests and experience in the area of transfer of training, including
those who conduct research in academic settings, those who conduct research
in Department of Defense laboratories, and those who both conduct research
and apply findings in non-academic settings. Further, the planning
committee invited a number of participants to comment on the presentations;
these individuals represented both military and non-military organizations.
A complete listing of workshop participants, including those who made
presentations, may be found on page iv.

Presenters were guided by a set of questions prepared by the planning committee for the workshop; those questions are provided on page vi. All presenters were provided with transcriptions of their remarks; the individual papers are the edited versions of the transcriptions. One presenter, Joe Yasutake of the Air Force Human Resources Laboratory at Lowry Air Force Base, did not prepare a paper for this document.

The summary and recommendations were prepared from concluding remarks made by all participants.

Special thanks are due to Gary M. Olson, who chaired the workshop, and to Cynthia H. Null, Executive Director of FBPCS, who provided much support in both the organizational and editorial tasks of the first Roundtable meeting.

Virginia E. Holt
Office of Scientific Affairs
American Psychological Association

PARTICIPANTS

Department of Defense Research Roundtable
February 27-28, 1986
Washington, D.C.

Earl Alluisi, Ph.D.
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas

Captain Paul Chatelier
Assistant for Training and Personnel Technology
Office of the Deputy Assistant Secretary for Research
    and Technology
Department of Defense

Susan Chipman, Ph.D.
Office of Naval Research

Stephen Cormier, Ph.D.
Army Research Institute for the Behavioral and Social
    Sciences

Stanley Deutsch, Ph.D.
Committee on Human Factors
National Academy of Sciences

Marshall Farr, Ph.D.
Arlington, Virginia

Virginia E. Holt
Office of Scientific Affairs
American Psychological Association

William Johnson, Ph.D.
Search Technology, Inc.
Norcross, Georgia

Milton S. Katz, Ph.D.
Army Research Institute for the Behavioral and Social
    Sciences

David Kieras, Ph.D.
Department of Technical Communication
College of Engineering
University of Michigan

Gerald J. Laabs, Ph.D.
Navy Personnel Research and Development Center

Alan M. Lesgold, Ph.D.
Learning Research and Development Center
University of Pittsburgh

Angelo Mirabella, Ph.D.
Office of the Deputy Assistant Secretary for Research
    and Technology
Department of Defense

William Montague, Ph.D.
Navy Personnel Research and Development Center

Cynthia H. Null, Ph.D.
Federation of Behavioral, Psychological and Cognitive
    Sciences

Gary M. Olson, Ph.D.
Cognitive Science and Machine Intelligence Laboratory
University of Michigan

Jesse Orlansky, Ph.D.
Institute for Defense Analysis
Alexandria, Virginia

James M. Royer, Ph.D.
Department of Psychology
University of Massachusetts, Amherst

Walter Schneider, Ph.D.
Learning Research and Development Center
University of Pittsburgh

Robert Swezey, Ph.D.
Behavioral Sciences Research Center
Science Applications, Inc.

Ruby Takanishi, Ph.D.
Office of Scientific Affairs
American Psychological Association

Paul Thayer, Ph.D.
Department of Psychology
North Carolina State University

Joe Yasutake, Ph.D.
Air Force Human Resources Laboratory
Lowry Air Force Base, Colorado

DoD Research Roundtable
Transfer of Training
February 27-28, 1986

Guiding Questions for Presenters

1. From your particular specialization, what can we look for in
    the future in the area of transfer of training? What
    approaches do you expect might be developed? With what
    results?

2. What should be taught (e.g., skills? strategies?)? How do we
    find out what should be taught? How should the amount of
    transfer of knowledge be measured?

3. How much should be taught, and how (e.g., classroom
    situation?)?

4. How can we ensure/measure transfer of training (through
    performance) on the job?

5. Is training enough? Is it worthwhile to train?

6. When do we train for certain skills? What skills should be
    emphasized in an initial training period? What skills should
    be included in later/advanced training periods?

7. What provisions should be made for feedback from the job to
    the training program?

8. How does personnel selection influence training programs?

9. How can the appropriate training programs be implemented?

William Johnson, Ph.D.

Search Technology, Inc.

My intention today is to speak about the reality of transfer of
training research. The presentation will be based on my experiences while
conducting transfer of training assessment in technical training
environments. I will first describe our transfer of training research then
will respond to the list of questions prepared by the APA and suggested as
our presentation content.

My background qualifies me to discuss the reality of transfer of
training assessment in technical environments, as I have a good deal of
practical experience as an instructor and as a student of technical
instruction. I taught in technical training environments including high
school, community college, and university for over ten years. I've been
involved in a number of technical classes in the Air Force, Navy, Army, and
industry where I've participated as a developer as well as a receiver (i.e.,
student) of instruction. In addition I have formally assessed transfer of
training in a number of technical environments.

My experience in transfer of training (since 1977) has been in
examining how a person transfers from simulation oriented computer-based
simulation to real equipment diagnostic/troubleshooting performance. It's
from that dual perspective — technical instructor and researcher — that I
approach these problems and make this presentation.

When considering a transfer of training experiment, one must think
about the importance of good planning, good communications, and the
importance of coping with problems that might arise in the various stages of
a transfer of training experimental design.

It is difficult to plan for all possibilities. One example is an
experiment associated with my dissertation. The transfer task required
troubleshooting operational aircraft engines running on an outdoor test pad.
The day before beginning the experiment, I came down with pneumonia, and for
the rest of the week received daily physician care before collecting my
data.

In that same experiment, we pilot tested our instruments first by
collecting the troubleshooting performance of a number of example subjects.
We assessed the pilot test data to be sure that our data collection
instruments were appropriate. We were reasonably sure that the instruments
would permit us to record and classify all of the actions of the subjects.
Our best laid plans met with difficulty by the second or third of about 40

1

subjects. With that particular failure the aircraft engine (transfer task) ran out of gas as planned. The subject didn't have the slightest idea what to do, therefore, put down the toolbox and cried. I didn't have anywhere on my form to check that the subject cried. In spite of the most detailed planning, it is impossible to anticipate everything!

My example of the importance of clear communication stems from a recent experiment in which we trained a group of unlicensed nuclear power plant operators. We used a computer simulation for training troubleshooting of an emergency diesel generator. The diesel generator is a standby electric power unit for a nuclear plant. We transferred people from either traditional classroom instruction (control group) or the computer simulation (experimental group) into the plant.

Of course, the person collecting the data needs to be blind to the treatment. Therefore, the classroom instructors told the students, "Whatever you do, don't tell that person in the plant what you did here in the training center." I was the person collecting the data. The first thing I said to each subject was: "It's important that you do not tell me what you did in the training center." The second subject immediately responded, "I won't tell you what I did, but I'll sure tell you what I didn't do and that was the computer simulation." Clearly this was a communication problem, therefore we quickly modified the instructions.

While planning and communication are critically important attributes for effective transfer of training experiments it is also necessary for the data collection personnel to be adaptable to unexpected situations. The first example of the importance of coping skills can also be highlighted by that situation where the technical troubleshooting subject cried rather than employing tools and test equipment. A second example of researcher adaptability is highlighted by two extreme examples of the data collection physical environment in which we have worked. We have taken data while working behind the cold blowing air of an airplane propeller in an Illinois winter. At the other extreme we have collected data in the nuclear power plant diesel room where the temperature was over 110 degrees F. When working in these environments we remind ourselves of the expression from graduate school that "data are sacred!"

I hope these brief stories serve to emphasize the importance of planning and communications as well as the need for the transfer of training researcher to have good coping skills for data collection. These attributes are particularly true when the transfer task involves "real-world" job performance. I would now like to briefly characterize our major transfer of training experiments. The most thorough review publications of this work can be found in Rouse and Hunt (1984) and Johnson (1987).

Table 1 characterizes our transfer of training experiments since 1977. We started with a computer simulation called TASK (Troubleshooting by Application of Structural Knowledge) that was context free (CF). Subjects were instructed to find failures within an abstract network. We varied such factors as the aiding in training, the nature of the feedback received, the number of problems received, and the complexity of the problems. In the next set of experiments, we used the same context free simulation and transferred people to a context specific (CS) simulation, name FAULT (Framework for Aiding the Understanding of Logical Troubleshooting). We

TABLE 1: TRANSFER EXPERIMENTS FOR STUDYING COMPUTER

SIMULATIONS FOR DIAGNOSTIC TRAINING

Training		Transfer	Years
Context Free (CF)	to	Context Free	77-78
Context Free (CF)	to	Context Specific (CS)	79
CF & CS	to	Aircraft powerplants	80-81
CS & real equipment	to	Army communications equipment	82-83
CS & real equipment	to	Marine propulsion systems	82&85
CS & real equipment	to	Nuclear plant safety systems	83-86

worked initially in the auto mechanics and the aviation training environments. As Table 1 shows, we later worked in the applied settings of Army Communications, Marine propulsion systems, and finally nuclear power plant safety systems.

One of our earliest findings was that problem solving can be taught with a context free simulation. That led us to become interested in how the learner might transfer that knowledge into real world problem solving. I will say more about these findings later in this presentation.

We also measured the training effect of job aids and ongoing feedback on post-training performance. We used a job aid to provide on-line "bookkeeping" to help the learner keep track of information during troubleshooting. The subjects trained with the on-line bookkeeping did better in the transfer task than groups trained without that aid.

We also studied the effects of various feedback during on-line diagnostic training. We found that feedback should explain when the person is making mistakes — not just typical kinds of feedback provided by some computer based instruction that make wierd noises when an error is committed. An example of the kind of feedback we provided is: "That is an incorrect and unnecessary step because your previous actions have already provided you with that information."

We found that a context specific focus was necessary to do relevant transfer assessment for "real world" troubleshooting job skills. I believe that if we train a person to do a job (e.g., troubleshoot a piece of technical equipment) the transfer research must look at the success of the learner transferring from the training environment to the real equipment.

We found that skilled technicians more readily benefit from generic diagnostic training than do novice technicians. The skilled technicians who were asked to troubleshoot an abstract network related the task to their real world troubleshooting tasks. The novice technicians said, "What's this? How does this help me learn to troubleshoot?" Consequently, the novices made more mistakes in the context free system. This finding suggests that novice trainees should initially be given simulation with a reasonably high level of physical fidelity.

In another example, students showed a preference for a rich feedback over "bells and whistles." As that line of research has evolved, the feedback has become far more sophisticated, based on the specific mistakes made by learners. As we keep raising the fidelity of the simulation, we have reached a point where the students are more attracted to the part of the simulation that provides rich feedback, rather than the higher fidelity part of the simulation. The bells and whistles approach high-fidelity, and in my opinion that's not necessarily where we need to go with diagnostic training simulations. Our research has emphasized the importance of cognitive fidelity over physical fidelity. A technical instructor at the Army Signal School at Fort Gordon told me that our simulations were "heads-on" rather than "hands-on."

During the remainder of my allotted time, I will answer the questions that were given to us as a guideline for our presentations.

Why do a transfer of training design? My response to that question is that transfer of training experimentation is a reliable means to assess the present training. It's a good method of comparing training methods, and it's also good for new device evaluation. If job training is the goal, it seems that transfer is a way of measuring whether you've accomplished that goal.

Unfortunately, transfer of training assessment is expensive, especially if on-the-job performance is measured. Planning, data collection and analysis, and report writing all require personnel — researchers and technical experts who must be paid. One potential solution is to allow or encourage the researcher to become something of a technical expert. It has saved a great deal of time in our research, even with equipment systems as diverse as telephone switchboards, diesel generators, or helicopter blade systems.

I think it's important to the quality of our research that we understand the technical environment, and I realize it's often difficult to do that. For example, if becoming a technical expert takes two years, it is not practical for the researcher to also become a technical expert. The team, however, must strive to obtain at least an ability to use and understand some of the jargon of the particular technical domain.

Another answer to why not do a transfer assessment when the organization, after seeing that the experimental treatment looks good, decides to train everyone with only the experimental treatment. Consequently, there is no control group for the traditional training, making it impossible to do a proper comparison. While such a situation does not preclude a transfer measure, it does eliminate the baseline measure provided by a control group.

Another problem is that training transfer experiments are sometimes just not required by management. They are not necessarily needed to see that transfer has occurred, so it saves a lot of time and money not to do it. Industrial training is usually chartered to train rather than do training research.

What are some of the logistics involved in a transfer of training experiment? The first task is to convince the organization of the value of the research. The plan needs to be translated into some language that makes sense to the people running the training organization. Potential cost savings, for example, is very motivational to training managers. The second task is to prepare a workable experimental design for the particular operations environment. You then design the experimental treatments, deliver the training, and collect and analyze the data. Although that may seem simple, so much can go wrong! Equipment becomes unavailable, instructional staff changes, and subjects become unavailable or "contaminated" by experience.

I think it's important to get the managers and instructors involved and psychologically "invested" in the project. However, interestingly, sometimes you can have too much organizational support. If everyone becomes too interested in the experimental training, like a computer simulation, it's pretty rough to run the control treatment as previously mentioned.

Two of the APA questions, e2 and e7, addressed topics related to teaching skill or strategy. If you plan to have strategy-based instruction, it is important to have some contextual focus. Just giving a person generic rules for troubleshooting doesn't work nearly as well as providing the general rules related to the real equipment familiar to the learner.

Further, the student needs a reason to learn. The student must perceive that the training is likely to reduce the uncertainty associated with on-the-job troubleshooting. They need to see that practicing in the training environment is going to somehow help them when they get out into the field. I think we need to build confidence through structured practice.

What do I see as the future of transfer assessment? I am interested in looking at the longevity of simulation devices — why one device seems to stay in the training program while others are used only as a research tool and then disappear. What are the implementation methods of certain new training, while one kind of training stays and another kind of training dies? While I do not intend to answer the question here, I must hint that part of the answer is found in how technical experts are involved in conceptualization design, implementation, and evaluation instruction.

Two additional areas that could affect transfer of training research are related to embedded training and expert systems. With embedded training, we ought to be able to collect data from the person while they're on the job without having a researcher there watching every move they make. The attention to student modeling in expert system/intelligent tutoring systems will also serve to enhance our understanding of learning transfer. It seems to me that these areas will make substantive contributions to the future of transfer of training research.

References

Johnson, W.B. (1987). Development and evaluation of simulation-oriented computer-based instruction for diagnostic training. In W.B. Rouse (Ed.), _Advances in man-machine systems research: Vol. 3._ Greenwich, CT: JAI Press, Inc.

Rouse, W.B., And Hunt, R.M. (1984). Human problem solving in fault diagnosis tasks. In W.B. Rouse (Ed.), _Advances in man-machine systems research: Vol. 1._ Greenwich, CT: JAI Press, Inc.

Stephen Cormier, Ph.D.

Army Research Institute for the Behavioral and Social Sciences

I would like to discuss the application of transfer of training research. I'd like to focus on DoD's MANPRINT program, because I think it poses some general problems in the application of transfer research and other human performance data.

MANPRINT is essentially a set of procedures and guidelines formulated to insure that the design of weapon systems and other complex hardware and software systems would be compatible with personnel and training resources.

In order to achieve transfer from training systems to weapon systems, it is essential to specify the human performance characteristics of projected systems. This goes beyond the usual consideration of human factors. Important information includes: the level of difficulty, the type of training to be administered, the minimum aptitude levels, and the approximate length of training needed to obtain satisfactory performance.

In order to be effective in guiding systems design, these types of training or performance analyses need to be provided while the system is on the drawing board.

The challenge for those of us who provide training design is to develop precise estimates of the training or performance implications of prototype or developmental systems.

The flip side of that challenge is the opportunity to have a major say at a point in the system when it's still possible to make a difference. Typically, we are presented with a fait accomplie — the system is already in production or operation, and we are asked to fix it.

Occasionally it is possible to fix a problem here or there in such situations, but often there are deeper problems. Of course, it is unlikely that an unworkable million dollar training system will be scrapped — particularly on the advice of a training psychologist.

There are two types of situations in the transfer world: the transfer of learning, when we're not quite sure what the transfer task is, and the transfer of training, where the target task is generally known with some precision. A good example of transfer of learning is the educational setting where people are simply being educated and they don't know precisely what job they're going to be doing. Industrial and military settings typically involve transfer of training.

7

The more complex the system, the more difficult it becomes to fully specify the nature of the performance task. The basic problem is that we have to try to estimate the nature of the task and develop a training task prior to the implementation of the system. Unfortunately, there seems to be an infinite number of transfer tasks that can be created for any given target task.

There are two other interrelated problems: basic research findings tend not to be applied effectively to real world situations and the applied setting is not necessarily the best place to investigate basic types of psychological processes due to the lack of experimental control.

I will present some of my own observations of the types of developments necessary for the provision of these types of training estimates.

First, I believe that the analysis of tasks should be based on the underlying cognitive and psychomotor processes involved in their performance rather than using the task procedures as our classification principle.

For example, I think it is very important to develop a consensus on the ways in which we can analyze tasks on ability dimensions. Can we come up with cognitive, information processing, and psychomotor factors that can then be used to analyze any given transfer task? Some of these factors might include short term memory, cuing conditions, and decision processes.

It seems to me to be important to get away from defining the task as a set of procedures; instead defining it as a particular way of tapping the information processing or psychomotor abilities of the subject. Once these analyses have been done, the training materials could be developed so as to simulate the task specific demands of those dimensions.

Second, the essence of being able to estimate or analyze tasks is to use these identified factors very specifically. For instance, in analyzing short term memory demands, are there portions of tasks where the person has to keep certain information in short term memory in order to be able to perform the task? How does the task structure affect the trainee's use of his/her short term memory, in other words. This is a more psychological way of doing job analysis than is customarily done.

If you are able to be this specific about task analysis, it seems it gives us much more power in the training environment. This specificity should allow us to be more analytical in identifying where the trainees are or are not having problems.

It is difficult to investigate the same types of issues in an applied setting that are studied in a basic research lab. I believe that applied research is more useful in testing the practical significance of factors already identified in the laboratory setting.

It would be useful to consider using different task dimensions in the applied environment, for example, repeated testing or extended intervals, which are often not done in the laboratory.

8

How do you verify the validities of the taxonomy or the consensus apart from determining that the consensus is a valid consensus? How do we know that we picked the right taxonomy or the right dimension?

The Services have spent a great deal of money in the last ten or fifteen years trying to develop tasks particularly based upon cognitive processes. They've been largely unsuccessful, particularly in reliably attributing certain underlying abilities to particular tasks.

For example, a number of years ago the American Institutes for Research attempted to develop some scales to describe underlying motor abilities, and to apply those scales to a particular task by using bench marks of one sort or another. The results generally were that judgments were not very reliable.

We want our models to allow us to specifically identify what is the effect on the individual of the training that he or she receives in terms of target task performance. For example, could we come up with an estimate of whether the occurence of task relevant stimuli would exceed the capacity for short term memory, given that they are not organized by differentiability, the characteristics of recoding from verbal to visual to auditory modalities, and so on?

Another example would be in the area of automatization. If we come up with a series of tasks, none of which could be automatized and which were being required to be simultaneously performed by the operator, we might predict that the operator would have difficulty dealing with the dual task situation unless one of those components could be automatized.

Some might suggest using the expert system methodology for the development of these models of cognitive or other abilities. Those individuals who are knowledgeable about the specific systems provide information that would be used for preparing the content of the training.

Clearly, expert systems can be useful, however, it must be done with great caution. There is potential for a loss of control in the development of the training model. How much control can we give subject matter experts in defining the basic cognitive abilities, for example? The history of such attempts suggests that in this complex area, task experts are not that reliable in their judgments.

The danger in trying to use the expert to lay out the instructional design is that their own understanding of how they are able to perform a task and which abilities they employ may be quite limited. As I see it, we have no alternative to the effective employment of the findings of basic experimental research on human abilities.

9

THIS PAGE INTENTIONALLY LEFT BLANK

David Kieras, Ph.D.

Department of Technical Communication
College of Engineering
University of Michigan

One thing I like about this meeting is that there's a mix of people here -- both academic and "real-world" researchers. I think I'm fated to represent the "unreal world" because I'm going to talk about those things that I view as helping to "push back the envelope" of what we know.

I have a particular position that I want to argue. I decided that I would not present some little tidbit of my research but instead try and show what I think a whole domain of research has to do with training or how training could be done.

I want to discuss the role of cognitive simulation in advanced training technology. The thesis is that cognitive simulation modeling is a way to obtain the specifications for what must be learned. This can then provide the input for intelligent tutoring and other advanced training systems and also low-tech systems, like paper-based training. Cognitive simulation modeling is not ready to be put into practice, but we should take it seriously as a potential approach.

I'm going to talk very briefly about cognitive simulation, in order to define it. Then I want to describe the role that it might have in training.

A cognitive simulation is a computer program that realizes a theoretical idea about mental structures and processes. The program contains explicit representations of mental processes and knowledge structures. State changes in the model are supposed to represent state changes in the mind at some appropriate level of analysis. The cognitive simulation efforts most commonly involve symbol manipulation programming, especially programming written in the LISP language using artificial intelligence (AI) programming techniques.

Why would one do such a thing? The classical reason is that this is a way to discover the cognitive processes, because this approach allows you to turn vague ideas into well-defined ideas. This approach also gives you the possibility of accounting quantitatively for complex empirical data. This is a relatively new application of simulation modeling.

As psychologists we have ideas that are often vague and ill-defined. These are ideas that guide our research and guide our thinking, but the scientific activity consists of trying to turn those vague ideas into

11

specific ideas. A simulation model is a tool we can use for that. It gives a lot of specificity and precision. The tradeoff is that they are difficult to construct and limited in application.

The simulation model, when properly done, is in some sense a realization of a specific set of psychological theoretical ideas. There's no point in building one of these models in the absence of good psychological ideas; in the initial use of simulation models this basis wasn't made clear.

This, the simulation model, is not in fact the psychological theory. That's a different set of ideas. But it's a more specific version of the psychological theories. By developing the simulation, you get some ideas about the consistency, the completeness, or the clarity of the theory, because you have to be specific and precise with a computer. You are forced to turn your ideas into a set of specific ideas.

Such simulation models currently are used to work with empirical data at levels of precision and detail that have not been possible in the traditional state of psychological theory. These days, with our laboratory computers, we can collect much more detailed data than was formerly possible. I believe these simulation models deal with data at that same level of detail.

Why would one do cognitive simulation? For several reasons: You want to clarify a theoretical idea by constructing an explicit realization of it. You want to demonstrate that the theoretical idea you have is sufficient to produce the behavior of interest. The history of psychology is populated with general statements about things that rarely were shown to actually be capable of producing the behavior that people were talking about.

Sufficiency is a question you can answer in a very straightforward way with a simulation model, because you can account for the behavioral data in detail. Once you've done that then the model provides you with this statement that summarizes all that data. So instead of looking at 20 graphs, you can say: "Here's the model and here's what it does."

What does a simulation have to do with training? I believe it is possible that these simulation models can provide theory-based predictions and evaluations of practical situations.

The linkage between theoretical ideas and the simulation program is best when the program is based on definite assumptions about the architecture of cognition — the basic structure of what's inside the head -- as opposed to just having code that produces behavior. This seems like an obvious point but it hasn't been made explicit before.

There is a cognitive architecture that's become very popular, and very important. A distinction is made between declarative knowledge, knowledge of facts, and procedural knowledge. Declarative knowledge is often represented in a semantic network, which is a way to express facts. Procedural knowledge is often represented by a set of production rules which examine and manipulate this declarative information.

What kind of information is in a cognitive simulation? In order to
specify a cognitive simulation model, several components must be present.
One is a detailed task analysis. The most important part of this analysis
is to clarify the goals this simulation model or the human must accomplish.

One also must be explicit about the declarative knowledge that's
required to do the task. Are there certain facts about the domain the model
has to have in order to be able to do the task? Likewise, what procedural
knowledge must the model have? What rules, procedures, and heuristics must
it have?

As an example, consider electronics troubleshooting. If we had a
simulation model of electronics troubleshooting, we would have analyzed the
logical structure of the troubleshooting task. We would have figured out
what were the important facts about electronics and the specific system
being tested, and we would have defined what were the various rules or
heuristics involved: for example, how do you infer states of components
from observations?

If the simulation model is successful, then we have some measure of
confidence that we have accurately and completely characterized the critical
knowledge. One cannot build a model for all possible tasks in a particular
domain. But once we have a model to do one thing, at least, we have pretty
complete characterization about that one task.

How can you use the information in a model? First, the task analysis
and knowledge specifications characterize the task itself. For example, you
could use it to evaluate the quality of the design of the system that the
trainee must interact with. This is in essence what Kieras and Polson
(1985) are doing in the field of human-computer interaction.

You can use it to compare jobs in terms of the underlying tasks because
you have identified the subtasks that are involved. So you're in a position
to determine which jobs have which shared tasks if you build enough of these
models. You could then look at the job and find out how much information a
particular job requires. For example, you might discover that a certain job
involves only a couple of subtasks being known and some other job might
require quite a few.

The specific contents of the declarative and procedural knowledge
provide an explicit specification of the minimum required knowledge in order
to do the task. From that specification you can tell what knowledge is
really important. For example, the classic problem in trying to train
people in electronics maintenance is an issue about what electronics theory
they really need. The experienced troubleshooter might say, "Ohms Law?
Yes, I learned that once and never use it." If we had that specification,
we'd be in a position to make much more intelligent guesses about what
knowledge people actually have to have to perform a task. We can look at
the courseware to see if it contains the right information. My intuition is
that often the training material doesn't have the information we need to
make a computer program able to do the task.

Finally, if you have these explicit representations, you have now
worked out a lot of the detail necessary to do things like build intelligent

tutoring systems. You have the characterization of what must be learned and you can export it to other kinds of systems.

The practicality of building these models presents a set of "puzzles."

First, can you build these models cost-effectively? In order to do that, we need some kind of technology of cognitive modeling; instead of an experienced cognitive psychologist fiddling with one for a year or two, a graduate student in cognitive psychology should be able to build one in less than a year because they would know what they had to do. Such technology requires a methodology for task analysis. If we could specify how to build one of these models from the ground up, we'd also learn a lot about how to do this cognitive task analysis we've been talking about. The modeling approach provides you with a way of verifying the task analysis on a small scale.

A second question is whether the really complex issues can be finessed. In some sense we shouldn't have to model everything about a task. We shouldn't have to construct an entire model of perception, for example, in order to construct a model of how people do electronic troubleshooting, because regardless of how the perceptual system works, it does deliver certain outputs to the rest of the system. We want to know what is in the rest of the system.

So a useful model would be one that had a lot of these perceptual issues that you could finesse, by saying the input to the system is the simpler output of the perceptual system.

But can we reliably identify and focus on just that critical information? Do we know how to finesse? Do we know how to choose what to finesse?

One of the more valuable things about this approach is that the information in the cognitive model is also the information we want in intelligent tutoring or an advanced testing system. How do we streamline the export of information from one of those efforts to the next? One of the things you have to do is work with compatible knowledge representations. It would be easier if there was a general consensus on the cognitive architecture we want to assume. Some examples of how this approach can be applied can be found in a recent technical report (Kieras, 1987, in press).

14

References

Kieras, D. E. (1987). The role of cognitive simulation models in the development of advanced training and testing systems. Technical Report No. 23, University of Michigan.

Kieras, D. E. (In press). The role of cognitive simulation models in the development of advanced training and testing systems. In Frederiksen, N., Glaser, R., Lesgold, A., and Shafto, M. (Eds.), Diagnostic Monitoring of Skill and Knowledge Acquisition. Hillsdale, N.J.: Erlbaum.

Kieras, D. E., & Polson, P. G. (1985). An approach to the formal analysis of user complexity. International Journal of Man-Machine Studies, 22, 365-394.

THIS PAGE INTENTIONALLY LEFT BLANK

James M. Royer, Ph.D.

Department of Psychology
University of Massachussetts

In my presentation I will describe a method for characterizing the
dimensions of transfer space and then talk about how that characterization
can be translated into a specification of transfer objectives that can serve
as the basis for training procedures. My ideas are speculative and I would
appreciate your thoughts on them.

There has been much discussion today about the various situations in
which transfer problems appear. My talk will be devoted to developing a way
of characterizing those situations and to identifying the kinds of problems
that one might conceptualize as transfer problems.

A Two Dimensional Representation of Transfer Space.

One way to characterize a transfer situation is to specify the
relationship between the physical similarity of the instructional task and
the transfer task. This is the old "identical elements" notion: the idea
that transfer is going to be determined by the degree of stimulus similarity
between the training task and the transfer task. In the discussions thus
far several speakers have spoken of the "fidelity" of training to job
performance; this is one kind of idea I am trying to capture in my
characterization of the identical elements dimension of transfer space.

Identical elements theory only takes you so far in characterizing
transfer tasks and I would like to suggest that there is a second dimension
of transfer tasks that is as important as the physical similarity dimension.
That dimension is the complexity of the knowledge that is being used to
perform the task.

In recent years two advances have given us better ways of
characterizing the knowledge dimension of transfer tasks. One advance is
the research on differences between expert and novice performance that has
emphasized the fact that knowledge differences between the expert and novice
are as much qualitative as they are quantitative. The second advance is the
utilization of production systems as the bases for theories of cognition.
Production systems provide concrete models of processing events that in
earlier theories were labeled boxes in a diagram.

My notion of characterizing transfer space using the dimensions of
stimulus similarity and knowledge complexity is represented in Figure 1. A
condition of training or instruction is represented at the far left

17

Figure 1
Description of Transfer Space

Knowledge Complexity Dimension

HIGH

LOW

0

Skilled Scientist

Novice learner

Creative Scientist

Skilled Technician

Low

High

Stimulus Similarity Dimension

18
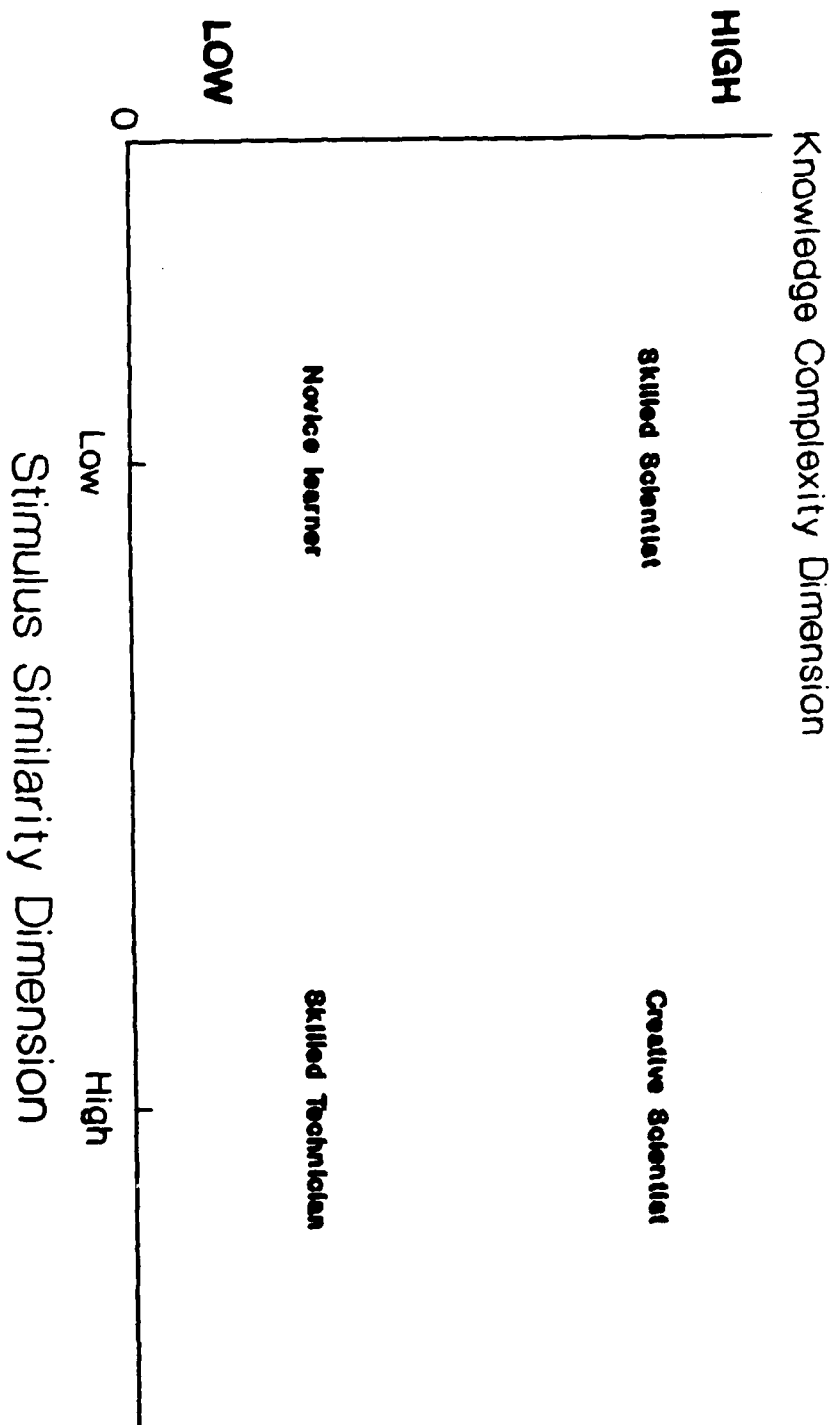
intercept of the graph in Figure 1. The horizontal axis of the graph represents the degree to which the transfer task is dissimilar (in a stimulus sense) from the conditions of original instruction. For example, close to the point of original instruction on the horizontal axis are tasks that closely resemble those that were considered during instruction or training. At the extreme right on the horizontal axis are tasks that can be completed utilizing the knowledge acquired in training but these tasks are quite different in a physical or stimulus sense from the conditions of original instruction.

The vertical axis of the graph represents the complexity of the knowledge that can be utilized to perform a particular task. The characterization is from very simple or low level knowledge to very complex high level knowledge. I want to emphasize that the notion of knowledge that I am trying to capture is not degree of knowledge, but rather kind of knowledge. What I have in mind is the sorts of differences that have been discovered between expert and novice physicists. The expert physicist appears to have his or her knowledge represented in hierarchical structures with the apex of the structure consisting of very high level principles that encompass an enormous amount of subsidiary information. That kind of high level knowledge is what I am trying to represent at the upper extreme of the vertical axis in Figure 1.
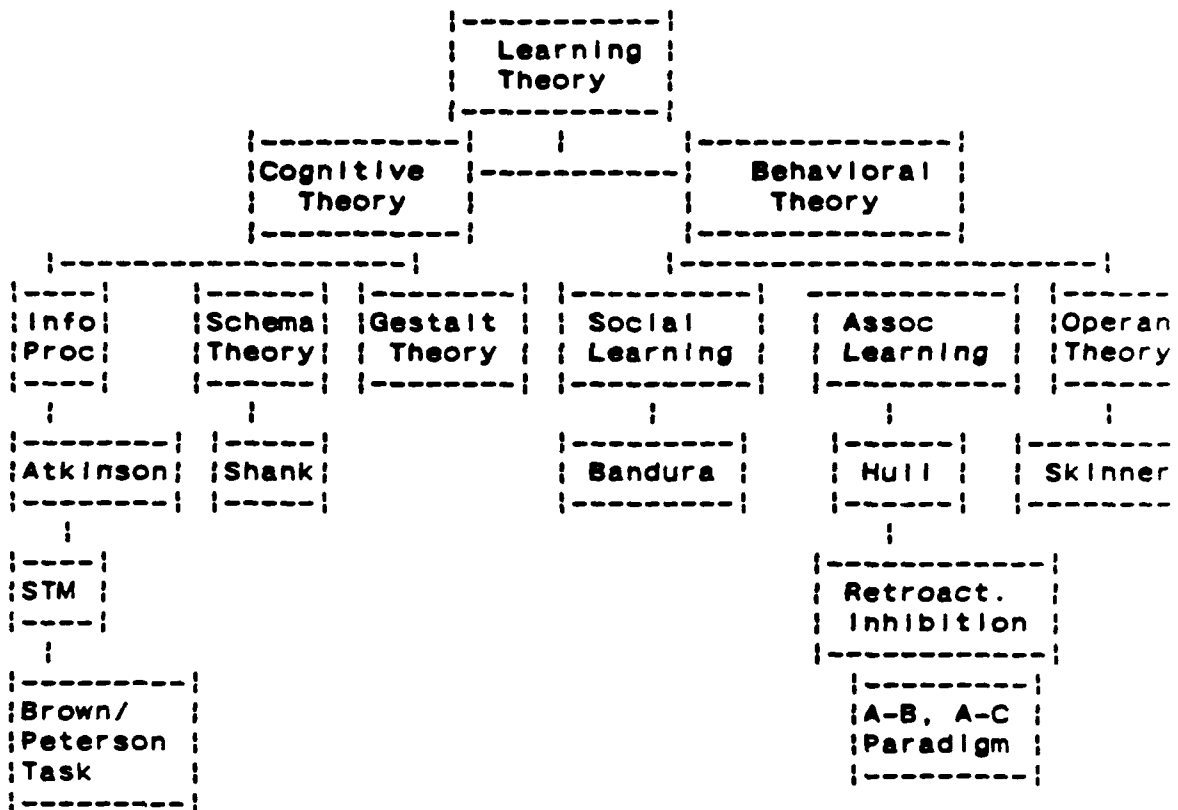
Within the graph I have provided some labels that might give you some idea of the kind of performance I am trying to capture with my two dimensions of transfer tasks. Within the lower left quadrant of the graph we have performance of tasks that require relatively low levels of knowledge and that are physically similar to those encountered in original instruction. In the lower right quadrant are tasks that require relatively low levels of knowledge but the transfer situation is quite different from the conditions of original instruction. In the upper left quadrant we find expert performance — the performance of tasks requiring the use of complex knowledge but still within the context of those conditions that existed during original learning or training. Finally, in the upper right quadrant we have the performance of tasks that require very high level knowledge in situations that are very dissimilar from the conditions that existed during original learning. This is the sort of performance that we might expect from people who make scientific breakthroughs.

Describing Knowledge Complexity

If the way of characterizing transfer tasks that I have described is to have any practical utility it must be possible to objectively describe levels of both dimensions. I would like to spend the next few minutes talking about how this might be done. I will begin with an example of how we might characterize knowledge about learning theory. It might go something like the representation I have depicted in Figure 2. As we go from the upper most node that I have called "learning theory" the hierarchy spreads into two nodes that I label "cognitive theory" and "behavioral theory." Under cognitive theory I branch into information processing theory, schema theory, and Gestalt theory. In behavioral theory I have social learning theory, operant theory, and associative learning theory. At the next level down I have specific theorists such as Atkinson, Shank, Hull, and Skinner. At a lower level yet I have specific phenomena and specific experimental results.

Figure 2

Knowledge Hierarchy for Learning Theory

```
                         |------------|
                         |  Learning  |
                         |  Theory    |
                         |------------|
          |------------|     |     |------------|
          |Cognitive   |-----------|  Behavioral|
          | Theory     |     |     |  Theory    |
          |------------|           |------------|
  |----------------------|           |------------------------|
|----- |  |------|  |---------|  |---------|  |-----------|  |-------|
|Info|    |Schema|  |Gestalt  |  |Social   |  | Assoc     |  |Operan |
|Proc|    |Theory|  | Theory  |  |Learning |  | Learning  |  |Theory |
|----- |  |------|  |---------|  |---------|  |-----------|  |-------|
    |         |                      |            |              |
|--------|  |-----|             |---------|  |------|     |--------|
|Atkinson|  |Shank|             | Bandura |  | Hull |     | Skinner|
|--------|  |-----|             |---------|  |------|     |--------|
    |                                            |
|----|                                     |-----------|
|STM |                                     | Retroact. |
|----|                                     | Inhibition|
    |                                       |-----------|
|---------|                                  |---------|
|Brown/   |                                  |A-B, A-C |
|Peterson |                                  |Paradigm |
|Task     |                                  |---------|
|---------|
```

I would not argue that the hierarchy I have depicted in Figure 2 is an agreed upon way of classifying this particular knowledge domain, but I would suggest that is possible for a group of experts to come to some agreement with respect to characterizing the knowledge hierarchy present in a narrowly constrained knowledge domain such as learning theory.

One aspect of the knowledge hierarchy as it relates to expert/novice performance is that when an expert is presented with some phenomena related to learning theory the expert activates very high level knowledge units automatically as a function of encountering phenomena. Some cognitive theorists have referred to this as the "proceduralization" of knowledge. I want to emphasize that the structure I have presented is not a model of how the content within a domain should be taught. I'm reasonably certain that teaching a novice all the labels and connections in a knowledge hierarchy would have little impact on the novice's ability to classify phenomena that were very different than those encountered previously.

The particular knowledge hierarchy that I represent in Figure 2 happens to be one associated with a cognitive domain. But it is also possible to do a similar kind of analysis for certain kinds of motor skills. For example, in tennis one can divide skills into shot skills and strategy skills. Shot skills are mastery of forehand, backhand, volley, etc. Strategy skills come into play in decisions about when to utilize particular shots. One of the distinctions between a truly expert tennis player and a good club player is that the club player may have good shots but he doesn't have the strategy components of the game. The expert player not only hits great shots but he hits the right shot at the right time; that's what percentage tennis is all about.

The thing that is common in my characterization of the knowledge hierarchies of the expert learning theorist and the expert tennis player is that a particular event in either domain automatically triggers off very high levels of knowledge in the domain. In contrast, the same event triggers off lower levels of knowledge in the novice.

Knowledge Complexity and Transfer

If one accepts the assumption that the same phenomena triggers off knowledge at different levels in the expert and the novice, this has implications for the characterization of transfer space that I presented in Figure 1. Consider people who master information at various levels on the knowledge level axis of Figure 1. Now imagine a range of transfer tasks the knowledge will transfer to. My hypothesis is that someone mastering knowledge at a relatively low level of knowledge will only be able to transfer that knowledge to a small range of transfer tasks that have a stimulus context that is similar to the conditions of original instruction. In contrast, the person who has mastered knowledge at a high level could be expected to transfer that knowledge to a much broader range of tasks. I have depicted this prediction in Figure 3. This suggestion leads to a further hypothesis. If you want to train for transfer you can do it in one of two ways: Expand the range of stimulus characteristics the trainee is exposed to during training, or extend the complexity of the knowledge the trainee is acquiring.

21

Figure 3
Description of Transfer Space

Knowledge Complexity Dimension

HIGH

LOW

0

Low

High

Stimulus Similarity Dimension
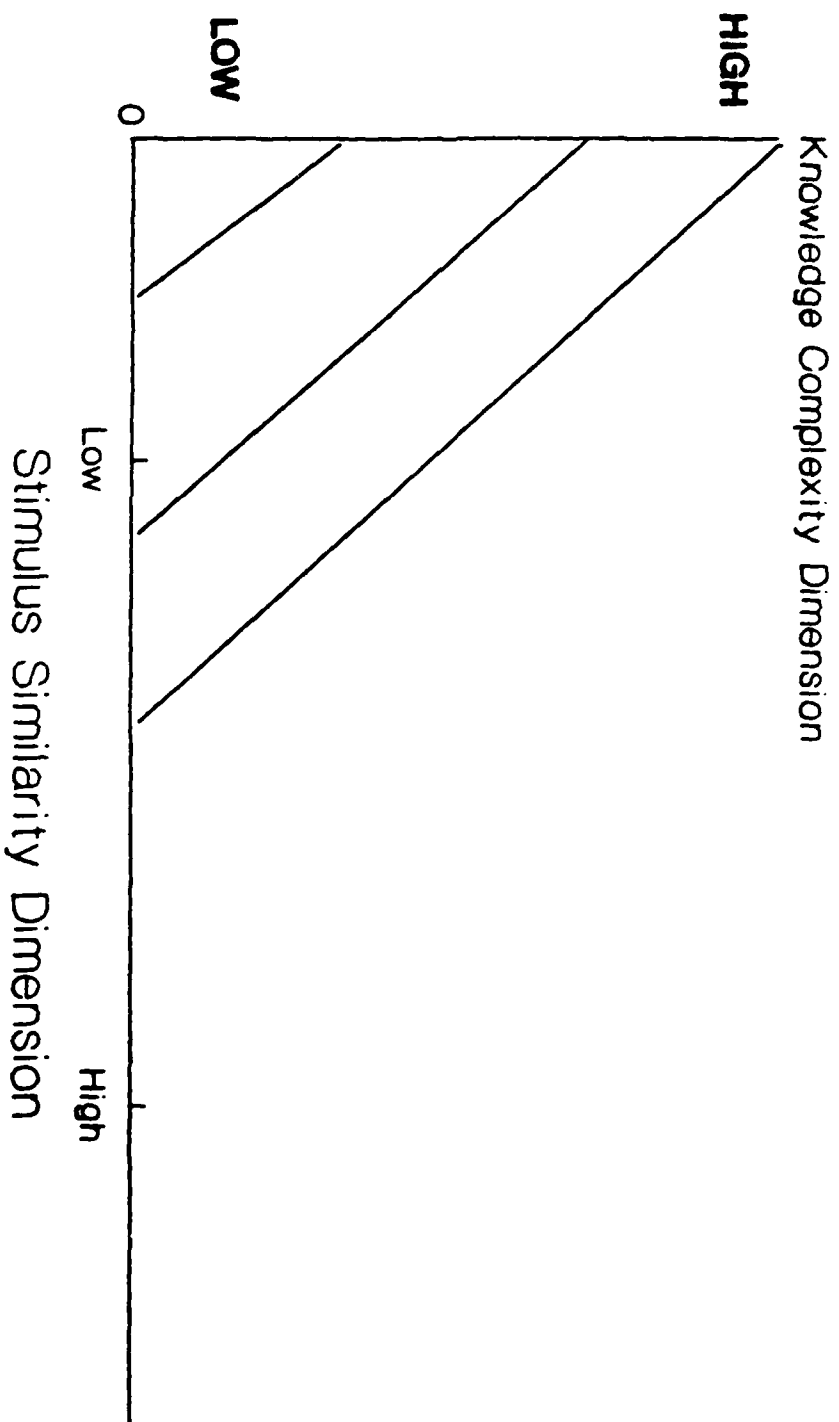
I envision the knowledge domains I am talking about to be narrowly constricted with a vertical rather than a horizontal architecture. That is, acquired knowledge within a domain can transfer downward to related situations, but information will not transfer across domains. For example, chess skills are unlikely to transfer to card or board games.

The two dimensional characterization of transfer space that I have described could be used in planning training that will transfer to other situations. The idea is that the instructional developers first must decide what trainees should be able to do after completing training. This description of what they should be able to do would include a description of the range of conditions the training should transfer to (you could think of these as transfer objectives). Let's examine several instances that we can represent on Figure 1. Suppose that an individual is being trained to repair field radios. Further, let's suppose that the radios the individual is being trained to repair all look pretty much the same and they have similar design and function. This situation could be represented in Figure 1 by imagining a transfer space very close to the point representing original instruction on the graph. An individual being trained to perform the function of repairing the field radios could probably perform his or her job successfully by being given a minimal amount of "textbook" training on electronic theory and troubleshooting and then practical training on actually repairing "prototype" field radios.

Now imagine a situation where an individual is being trained to repair a wide variety of electronic equipment. In other words, the training is expected to transfer to situations on the far right of Figure 1. Now training can either consist of relatively low level abstract learning and considerable practical experience on a wide variety of equipment (i.e., expanding the physical similarity dimension of training), or it can consist of focusing on higher levels of knowledge that will also result in the expansion of the range of problems the trainee can successfully handle.

It seems to me that the trick in carrying out the kind of analysis I am describing is determining exactly what you would like a trainee to do. There are probably many situations in military training where trainees are only going to be faced with job situations that can be virtually duplicated in the training environment. In this kind of case low level knowledge training coupled with practical experience would probably accomplish the training goal. In other situations, however, it may be very difficult to specify the range of tasks the trainee will be required to accomplish in the job situation. In this case the trainee should be prepared for the maximum transfer potential which means the acquisition of complex knowledge accompanied by experience with a range of tasks. A further complication associated with specifying the range of situations that one would expect training to transfer to is that job requirements change over time. The individual who has been trained to perform one kind of task on a limited range of equipment may, over time, find that the requirements of the job have changed and the initially received training is no longer adequate for job performance.

Summary and Conclusions

In my talk today I have presented a means of describing transfer space in terms of two dimensions: a dimension describing the physical similarity

between the conditions present during original instruction and those present in the transfer task, and a dimension describing the complexity of the knowledge that is relevant to completion of the transfer task. This description of transfer space could be useful in formulating "transfer objectives." Transfer objectives are instructional objectives that describe the range of tasks that training or instruction is supposed to transfer to. I also pointed out several implications and predictions associated with my characterization of transfer space. As an instance, I suggested the hypothesis that mastery of knowledge at relatively high levels of complexity could be expected to transfer to a broader range of transfer situations than would mastery of lower level knowledge.

William Montague, Ph.D.

Navy Personnel Research and Development Center

I want to contrast practical notions of transfer with
laboratory/research notions. This is a contrast between schooling and what
I'll call "mentoring".

"Schooling" is the experience that most psychologists have that gives
us a particular mental model that we use to deal with the world.

Schooling is a situation with a general goal — to build an educated
person through elementary school, secondary school, and other schooling.
Schooling is not tied into practical objectives for the most part, and in
fact, if you listen to educators argue about curriculum and curriculum
planning, you find that the arguments focus on how to specify the generic
goals of education. The goals are quite abstract. Yet, because we worry
about transfer, we look at these abstractions as if we can do something
about them. For example, "can we train generic problem-solving skills"?
Can we train the components of reading if there are these components, like
decoding skills and certain vocabulary access? If we train those skills,
will they transfer and make everybody better?

Unfortunately, generic training programs don't seem to work well, and
because of the level of abstraction, I would argue that there is
considerable evidence that shows a lot of this is difficult for students.

It's difficult for students to learn abstractions (e.g., generic rules
for solving problems) that are separate from concrete examples. Research in
cognitive psychology has shown that you need concrete examples to facilitate
learning.

Many of the abstract things that people are supposed to learn, in
geometry for example, are not made explicit in the instructional situation.
They are always to be discovered, to be uncovered, and, of course, most
students never do uncover them.

There's a need for explicitness. The science education literature
provides evidence that there's weak or little transfer between the academic
abstractions learned in class and practical situations where you have to
transfer (use) that knowledge to solve problems.

DiSessa demonstrated fairly conclusively with physics that what we get
out of the educational, instructional, or training process is a separate
body of academic knowledge that applies in class but not in situations where

25

that knowledge should transfer. The notion is that information doesn't transfer because the connections between the abstract formalisms and concrete events are not made properly.

Most psychologists and educators have the viewpoint that this schooling model is the kind of model we expect in a learning, training, or the instructional situation. This influences two things: It influences what research they do and the way in which they begin to think about transfer and transfer of training; and it influences the instruction they design and, I think, fosters the lack of attention to the strategic processes that are part of the instructional process.

If we look at the "mentoring" situation — or perhaps a better term would be "experiential" situation — we're really talking about a vocational or apprenticeship model. It is involved with domain specific performance. The result of the instruction process is to build a domain theory into the individual, to guide his or her performance. The only time you need to teach the individual about formal theory is when that theory is directly relevant to what he or she is going to undertake. The practice given is concrete and very specific. It is usually also implicit that it's a one-on-one instructional situation. In that situation strong transfer is a primary objective but it is of the kind that's been called "near" on a "near-to-far" transfer dimension.

We need to do a more adequate analysis of the strategic, metaphorical and analogical knowledge needed for far transfer.

I want to contrast teaching abstracted components of skill with simulation, because from the practical viewpoint, if I need to train people to do particular jobs, the primary emphasis will be to mimic situation characteristics, or use simulations, of the actual job.

If I'm going to do that in school, I'm going to push toward the use of simulation and role playing. I don't care whether it's paper and pencil simulation or computer-based simulation, but the emphasis is on how I can represent critical functional aspects of the real task.

Up front is the issue of realistic simulation, not analysis into psychological components. When we talk about generic notions of transfer, we get involved across domains and require abstract conceptions about the skill that transfers.

There is confusion between the specific aspects of tasks and the strategic knowledge about how you act in this particular situation, i.e., general approaches to the world. For instance, your chess skill/knowledge might apply to "Go" or checkers.

Notice that in crossing domains the focus is on generic processes. When you start with a job-specific viewpoint, the focus has to be on representing that domain to a student, not teaching abstract rules.

One of the lessons the people from the "schooling camp" might consider is how to describe, analyze, represent and model the "jobs" that students might be expected to do. Deriving instructional situations from these descriptions is simpler.

26

The military has adopted a system planning model or procedure for attempting to relate training to the job. You derive the training practices, the descriptions of what you mean by competence, and how you will go about testing all from analysis and representation of job tasks. That's the theory, and the way the process is set up. It is a sound set of ideas and can be effective.

This procedure is not implemented well. People adopt an academic model of schooling rather than a mentoring model. Let me illustrate the point.

We looked at a course that was part of the propulsion-engineering curriculum. The course is supposed to teach the student how to do a chemical analysis of water that comes out of the condensers in the boiler system.

You would expect the curriculum to include some very simple basic chemistry and procedures. Instead, they have a full two-week course in chemistry — probably far more than is necessary. Again, the schooling model imposes itself on the training developers and they then don't follow the mentor/experiential model very well.

I have some recommendations for research and development. First, there is a history of 20 or 30 years of doing task analysis systematically in the military and the industrial world. That process is only partially adequate, because the strategic knowledge and the organization aspect of the task are not in that model. There should be research to improve that process.

In all instructional situations we allude to the entering student's knowledge. We use selection tests and perhaps some specific knowledge tests, but none of these are necessarily systematically related to the course. If it is true that a student learns well on the basis of his old knowledge, which is what many cognitive scientists would have us believe, then it is important to know what that "old knowledge" is to adapt instruction appropriately.

There is a problem of instructional fidelity. How do we take apart the critical aspects of the task to identify what pieces (components) of the task should be the ones that get practiced during the schooling?

Performance measurement takes too much time and energy to be practical. Therefore, if I ever want to measure competence or even expertise, how do I measure its level? In chess and similar domains, there are competitive systems where people have to deal with one another to get master points. That's how they are rated one against the other. How do we do that in other situations without taking lots of time?

I think it is worth emphasizing that the people who are going to use this knowledge are not the people in this room. Instructors and teachers in the real world are the ones who have to apply what results from better systemization. How do we reach them? We need automated implementation aids to help with the analysis, the design process, and the design of evaluation systems. We need automated systems that include cognitive and psychological knowledge.

27

Finally, I think that implementation of systematic instructional development processes must be improved. A feedback and monitoring system for the whole process is essential. The model I have for that feedback and monitoring system was described by Bill Cooley in his Presidential address at an annual meeting of the American Educational Research Association a couple of years ago, where he talked about a model he built for the Pittsburgh school system. Cooley's model takes measures out of the school system, identifies a problem, and then institutes changes with teams of individuals who would solve the problems via appropriate actions. I think the same sort of thing can be established for other situations.

The Navy has a feedback system in which job supervisors of recent graduates complete a questionnaire that is related to job objectives. This system doesn't work very well and is incomplete, but it does identify certain real problems.

The questionnaire asks if the student, when he first arrived from school, could do a certain task. It is a partial system — just a survey. It does not have a feedback loop in it to identify a problem and tie it closely to the training process. These supervisors are reporting on students several months after they've arrived, but they're supposed to reference them in time to when they arrived, because you don't want to confound it with whatever they've learned in the job situation.

In the military in a peace time world, relatively few people with a poor academic background are recruited. Under peace time pressures to keep expenditures down, selection of higher-level people is cost-effective.

If we get into a wartime mobilization, all of a sudden we accept lots of people who we would screen out right now; training those people is more difficult. How can we train these people more quickly?

Many issues of transfer have to focus on that relatively low end — I don't like to use the word "intelligence," but I don't know what else to use for that dimension. Work was done on this problem during the second World War, the Korean War, and more recently in "Project 100,000," when 100,000 people were taken and the military had to worry about whether they could utilize them in technical areas for which they were at least nominally unqualified. From recent analyses they performed nearly as well as higher ability people.

Military training is always going to be much less generic than ordinary education. For example, the Navy has Military Occupational Specialties (MOS), which are defined as a group of job sub-tasks. Training includes a general basic course, then more specific course(s) which tend to be very equipment oriented. The fundamentals course is supposed to provide the basic knowledge and skills required for that job. Even that is an academic model, but it is tied much more closely to the job in the design of the curriculum.

For example, the fire control technician, who is involved in shooting missiles, and fixing those systems that shoot missiles, is going to learn about fire control and about particular systems before working on these systems.

28

It is the fundamental (first) course that becomes like other academic courses. The fact that these generic courses are more "schooling" is of concern. They should be briefer and more job-like so that the trainee can fix the gear and be useful right away. "Experiential" courses would be better, if we could develop them. In this model, the student is given job-like experience and is brought to the job as an apprentice to help him pick up the fine arts. He has a period of time on the job with much less well controlled supervision that would be the case in his schooling. That's where you would begin to see the transfer of training — in his ability to learn on the job. Current training is too academic and does not prepare him for learning on-the-job well.

Because of marginal training, he doesn't get a chance to refresh himself on some of the things that were explicitly included in his training courses. Often, the guy who most needs training will spend the day sweeping floors, because they can't afford to risk practice on the equipment. Many people come out of basic training courses without having acquired the "basic" skills. For example, there's not enough practice soldering, not enough practice using test equipment, and no systematic training in how to find intermittent connections in cables.

Experiential models are difficult to implement, due at least in part to resource constraints. The model, although an old one, is still good. In fact, Prosser and Quigley published a book in 1925 suggesting such principles for teaching. It was more cognitive than behavioristic at that time. The environmental context and the situational context variables were very important for the instructional process. You had to represent to students the real situation insofar as possible and give them experience functionally equivalent to the job.

If we have certain tools for developing instruction, we'd have a better chance of implementing these models. For instance, we are working on an under-funded project on building an intelligent authoring system. The idea is to provide an environment in which someone who is beginning development of instruction — even in the analytic process — would be aided and advised by this system. You can think of it as an advisor or coach, but it would also help carry out the mundane tasks, such as the generation of textual materials and graphics.

It would also ultimately have a rapid prototyping system so that the individual could see the representation that he wanted to show to his student and try it out, and if he didn't like it, he could throw it away and develop a new one quickly. You can't do that kind of thing very quickly today, and people like Jim Hollan at NPRDC and Doug Towne at USC and others are working to build devices that are pieces of this ultimate system.

One issue in developing prototype simulations for training is what I call "instructional fidelity." People who develop simulators emphasize fidelity to the task and equipment actually used. Such devices are costly and may not be the best means of training because, in representing the task in all its complexity, it is inefficient for student learning. Instructional fidelity is concerned with how to implement, for example, part-task training, backward chaining, time compressed practice. Such procedures are needed to make instruction more effective and efficient.

Veridical simulators or real equipment can't do that. Thus, instructional fidelity issues are important for transfer.

Alan Lesgold, Ph.D

Learning Research and Development Center
University of Pittsburgh

I want to relate some experiences that my colleagues and I have had in looking at several specific technical areas, and in trying to determine how training might be improved in these situations.

Second, I want to say a few things about the kinds of knowledge that I think are needed to drive intelligent training systems or expert training systems. Finally, I want to comment on certain misconceptions and the role they might play in thinking about transfer.

We'll start by looking at the training of avionics technicians in the Air Force.

There's a kind of "high-tech" trap that exists in training these days. People come along with products to sell to the DoD, or people within the DoD who want particular products shape their definition in ways that make the products look as good as possible.

Inevitably, somewhere along the line we come to believe that the devices we bought are really simple — so simple that anyone could fix or use them. That effect is one of the reasons why certain kinds of specific procedural training find their way into the training process.

For many systems there are books that purport to describe exactly what to do in every circumstance, so that the real task is to train people to follow the instructions, to do so quickly, and without being too original in the implementation of those instructions.

The problem is that it doesn't work that way. The systems are never as foolproof as we might have hoped, and that in turn produces a problem. We train a person to follow a very specific instruction, and suddenly we realize he has to do some additional tasks just to survive. But, he was taught only instruction-following, rather than some more flexible knowledge that might have helped him deal with complex cases.

The example I want to cite involved people who used a relatively elaborate test station to fix the little boxes of electronic stuff that they stick in planes. The stuff in those little boxes is very reliable, but the test stations have been around a long time and are not as reliable.

It is important to note that planes that are used in the Air Force have been designed relatively recently. They undergo a certain amount of

31

modification over time. Often, when plans for new planes are made, a
reliable test station is included; but as budgetary pressures arise, the new
test station often disappears.

So, some current test stations are 1950's-level discrete logic. They
are unreliable things, but the premise on which the training of those people
is based is that the test station and the book of instructions will allow
them to fix the boxes in the planes.

When the test station works, that idea works very well. When the test
station fails, however, the idea fails badly — there are on the order of 40
cubic feet of discrete logic described by books full of schematics that have
to be used by somebody not well trained in electronics troubleshooting, but
only in following directions.

That's the kind of problem we have examined. We have compared people
who were recently trained and were doing well on the job with other recent
trainees who were not doing as well.

If we talked to first-line supervisors and were careful about how we
asked questions, we could get a reasonably good split of people according to
their competence. We could then do a contrastive analysis, look at the best
people, compare them with the worst people, and look at the differences.

We gave them a battery of tests, including some realistic trouble-
shooting tasks that were more difficult than they characteristically would
find on the job. But the original split is essentially asking the
supervisor who they think is more effective.

It is important to note that it's a face-valid split — people in one
group can do the work we'd like them to do, which is to troubleshoot, and
the people in the other group are much less able to do that. We have a
reasonably good ability to split the guys who are moving along quickly from
the guys who are moving along less quickly.

Now, it could be that there's some quality that some people come with
and other people don't, and the people who have that quality learn on the
job more quickly than other people, and all that we're doing is placing
people at different points in the course of learning on the job. Even if
that's the case, it's useful to know what it is that people are slow at
learning, because that's a useful focus point for more specific instruction.

What do we know about the more successful people? The high-skill
people are better at troubleshooting. They are more likely to find out
what's wrong with the things that they're trying to fix. They know specific
things to do and strategies to follow that are useful for the purpose of
troubleshooting this particular equipment.

We can also ask such questions as, "Are they more planful?" "Are they
more systematic?" "Do they have divide-and-conquer strategies?" There are
no differences in general strategic capability per se. The high-skill group
knows more about the systems that they work with and the specific components
that play a role in troubleshooting decisions. They know how those
components work. They know how they relate to the system as a whole.

32

They aren't any better at telling you how a resistor works. They
don't have specific electronics knowledge, and again, that's not bad. These
people are doing their job — things are going moderately well. Some of
them are doing their job quite well, and they still don't show a great
knowledge of electrical concepts.

And finally, if we look at the simplest level of procedures, we didn't
find any difference in their accuracy in doing things, like checking to see
whether a particular meter reading fell within a space given.

There were some more complicated differences of a similar character.
For example, if we look at how reliable our people are at making meter
readings, they're all about the same. If we ask the question, "Did they
make meter readings when they should have," then we find some interesting
differences.

Particularly, we find that when resistance needs to be measured in
order to diagnose, they were all about equally good, and they all tend to do
it. However, when voltage should be measured, the less skilled group was
much less likely to take the measurement at the appropriate time. Measuring
voltage requires deeper understanding than measuring resistance.

Our diagnosis measures are measures of the performance which they have
been hired to do on the job. When we say that they are better at diagnosis,
that is their job.

We split on supervisory ratings, but the truth of the matter is that
you always get curious and you go back and ask, "What if I split them on the
basis of their performance in the realistic troubleshooting task," and you
get much the same kind of result.

The difference between the two groups is that the domain-specific
quality of the strategy varies.

For example, suppose I have a computer in front of me that has ten
boards in it, and somewhere else in the shop there's a computer that works
and also has those ten boards in it. If I were systematic in my approach to
fixing the broken computer, I would take out board number one, put a good
board in, and see if that does it. If that doesn't work, go to board number
two, and so on. That's a perfectly systematic strategy, but it fails to
take account of the fact that the computer is more than the sum of those
boards.

Further, it fails to take account of the possible interactions. If
board number ten is bad, and is bad in the way that will cause board number
one to burn out, I can play the swapping game a long time. I once saw a
technician replace the same board seven times and he smoked each one.
People can behave very systematically and just not be doing the right thing.

If you look at the question the way we posed it, I think you would say
that we found higher skill people have more strong methods that would
generalize. It wasn't our intention to make that cut, but if you look at
what we're calling strong methods, there are things like tracing outs,
following schematics, and making certain kinds of measurements with the
meter.

These functions are not, in fact, specifics for one particular test station. In fact, it's the less-skilled person that says, "Gee, the only time I saw one of those fail, it was because board number 43 was bad. I'm going to try board number 43."

If you look at what the less-skilled persons were taught compared to what the high-skill people seem to be better at, a certain kind of generalization can be made. In the declarative knowledge arena, if you're looking at factual knowledge concepts of electronics, they were taught extremely abstract concepts at the very beginning of their careers.

What they don't get at the beginning is the more concrete stuff that has the character of the chapter in any reference manual called "theory of operation." If you're trying to narrow the problem to one of ten boxes in this big machine, then knowing what those ten boxes do can be very important.

Knowing the function of resistors may or may not be as important for that piece of the decision. One thing that seems to be the case is that the declarative knowledge that's provided is too abstract either to foster easy learning or to enable good troubleshooting.

On the other hand, the procedural knowledge that is provided in training seems to be too concrete, and characteristically involves the practice of merely following the instructions that are in the technical orders.

It is a replicable and predictable accident that troubleshooting jobs tend to involve more real problem solving than the system designer anticipated.

Now, let me try to deal with the second point. It is interesting to look at what really happens when we try to build a learning hierarchy or ordered characterization of the goals of instruction.

There are many different viewpoints one can take on how to split up the big connected fabric of knowledge that we want people to develop. In the case of some simple electronics, we can categorize by which concepts are being used; by kinds of circuits, parallel and series; and by kinds of laws. All of those splits tend to represent different viewpoints on the same basic task — we have a big body of highly interconnected knowledge that we want to convey to somebody.

How we make those splits though, can cause certain things to happen. Because of the need for accountability and the need for modularity in instruction, we split the pie into some pieces. Different people might be in charge of teaching those different pieces. Each of those people is going to want to verify that they have, in fact, correctly and completely done what they were assigned to do. So, they're going to test whether or not they have taught the piece we gave them to teach, and they're going to direct all of their attention to optimizing performance in those tests.

For example, let's say there are three things in particular that are useful to know in troubleshooting these big test stations. When we build our learning hierarchy, as we accumulate subskills into certain higher level skills, there's always some material left out. Some information is implicit

34

and we hope the student will figure out on his own — some glue that takes these separate pieces of knowledge and hooks them together to make them useable.

If I am in charge of teaching Ohm's Law, and I look at all the things I know about Ohm's Law, what am I going to give as a test to prove that I really taught the guy Ohm's Law? What am I going to teach him? I'm going to find the very core of that Ohm's Law knowledge — the thing that looks more like Ohm's Law than anything else, so that when people get that test item correct, I can smile and say, "I did my part." I'm going to focus my attention on the center of this piece of the fabric of knowledge we want to convey.

When we're finished, we have a situation in which the Ohm's Law teacher taught the central things about Ohm's Law, and the divide-and-conquer strategy teacher taught some of the central things about divide-and-conquer strategy. Each of them can demonstrate conclusively with their tests that what they set out to teach, they in fact taught. But there's a big assumption that the inter-relationships between those pieces of knowledge, will somehow, be successfully acquired by the student.

What do I do if I can't teach the student to troubleshoot the test stations, after he has taken these course modules? I complain about the fact that he doesn't know Ohm's Law and meter reading, because he can't use them to do what I want him to do. I send him back to those teachers and they teach again the same thing that they probably successfully taught before. This is insufficient. Different viewpoints are needed on the target knowledge for instruction, for remediation, and for assessing transfer.

If, for example, we want to have somebody know those aspects of Ohm's Law that are critical to troubleshooting, we need to identify and understand those aspects. When we find that people aren't learning to troubleshoot, we need to be sophisticated enough to understand that this is not a simple question of needing two weeks instead of one week's study of Ohm's Law. Rather, there is some other material that needs to be added that wasn't the target of instruction in the first place.

Let me present an example to illustrate my third point. The example is from Herman Hartel, who looks at children in the top end of the gymnasium in Germany. They get several years of physics and you can convince yourself that these students really understand the basic concepts, like Kirchhoff's Law, that help up apply some of the things that we know about electricity.

They can pass any sort of standard test that you could pose to those people. Then Hartel gives them a problem using real devices — wires, bulbs, and batteries. He says, "Look, here is this battery and the wire goes from here to this lamp and then over the next lamp, and then it goes back there, and this lamp lights up and this lamp doesn't. What's going on?"

And what do the students say? They say something like, "Well, the electricity gets through this bulb and then it stops here because this lamp is broken."

Tested in any conventional way, they will behave as if they know all the laws and principles of electricity that says "this simply cannot happen."

The point is that we would hope these really bright students would simply realize that this is a series circuit, and if the circuit is closed, which is the only way this lamp can light, the current has to be passing through that lamp. Whatever solution I come up with, it has to have that aspect. Instead, they retreat to a fundamental misconception — that electricity is the stuff that you push through wires, like electricity hoses, and it's quite possible for it to pass through this lamp and light it and then stop at the second light.

The point I want to make is that instruction in areas where there are these fundamental misconceptions has the character of war. It's a war against misconception, and when you fight a war there are certain basic principles that need to be kept in mind. One of those principles is that a particular amount of force will produce a victory only if used properly. At times you need to use overwhelming force in order to absolutely demoralize the enemy.

In this sense misconceptions that persevere represent fundamental conflicts with the skills that we want people to acquire, and they have to be shattered. That is the classic approach.

There are other things we can do that work better if we don't have the resources to try the overwhelming force approach. For example, one approach is the ancient tactic of building a series of garrisons. We have to stake out pieces of territory in order to retain control. That is a relatively expensive strategy.

In the case of conducting a war on misconception, that would mean having to provide enough pockets of knowledge to sufficiently cover the domain in question. In this way, we can count on the misconception being unable to reassert its control.

The third strategy is to win the hearts and minds of the people. You can provide some infrastructure that's there on your side even when you are absent. We do that when we teach certain kinds of metacognitive skills — when we convince people that certain constraints must hold in a domain and that they have to reason their way around the apparent contradictions to those constraints.

I would assert that when we think about transfer, we must keep in mind the possibility that transfer is often to tasks to which the person does have relevant knowledge, but some of that relevant knowledge is wrong. In electronics, for instance, when we deal with certain mechanical kinds of situations, people are equipped with a supply of successful principles that don't generalize.

Those principles have been reinforced in the average recruit by 18 years of experience in the world. Even if we taught a lot of the right things, the transfer situation — unless it's been planned for — will simply fall into one of those holes, which brings about the old misconceptions.

We can work with an expert ahead of time and go into a training situation knowing in advance almost every response — both correct and incorrect — that's going to be made in a very complex troubleshooting task. That means we can get to the point of knowing where to head off inappropriate experiences.

A final point I would like to make is that I think psychologists grossly underestimate the extent to which they need to be knowledgeable in the domain in which they are working. If you look at the first year or two of our work, it suffered greatly from being good psychology and superficial electronics. It was only after everyone on the project attended regular electronics seminars that we began doing really useful work.

THIS PAGE INTENTIONALLY LEFT BLANK

G.J. Laabs, Ph.D.

Navy Personnel Research and Development Center

I want to raise some issues today related to transfer of training that
have to do with applied performance testing. They are in the areas of
fidelity of measurement and the process by which we define what it is that
we want to measure.

I have spent the past ten years at the Navy Personnel Research and
Development Center doing research involving applied performance testing.
Initially my work involved building diagnostic testing and shipboard
training packages. This was followed by work on proficiency assessment to
provide feedback to the training community.

About three years ago, I switched from working in the training area to
the manpower and personnel area. Thus, my comments today will reflect more
of the viewpoint that would be taken by a personnel psychologist than a
training specialist.

The first issue I want to raise concerns the fidelity of performance
measurement. Applied performance testing, conducted at the appropriate
measurement levels, may be the best way to measure and ensure transfer of
training. By applied performance testing, I mean testing the individual on
the tasks that they actually do on the job using the equipment that they use
on the job.

This is an extremely high fidelity type of measure — the type of
measure that almost everyone would agree that you should use to assess
whether there is any transfer of training to the job. Unfortunately, this
type of high fidelity test is extremely expensive.

One of the major contributions of work now being done in applied
performance testing will be the development of a technology for building and
evaluating hands-on job performance tests — and perhaps other measures that
can be used in place of these job sample types of tests.

We have the complete technology to build paper-and-pencil tests, and we
also have a solid foundation for the technology to build criterion-
referenced tests, but we don't have the same technology for constructing
hands-on performance tests.

This lack of technology has implications for transfer of training work
as well as for the type of validation work I do now. For example, we don't
have any good technical guidance for selecting transfer tasks to properly
assess transfer of training or for selecting tasks for a test to validate

selection instruments. Clearly, performance testing is an area in need of additional research.

It seems to me that it is important in pursuing future research to develop some high-fidelity but low-cost performance measures, whether our purpose is to assess transfer of training or validate predictor tests.

It has already been mentioned that hands-on performance testing is expensive. It is expensive to build the test, and then very expensive to administer the test in a one-on-one situation. In addition, we take up the resources of the community in which we work — in equipment use, for example. If we use a less expensive, lower fidelity test, however, we do not know what effect the use of such a test will have on performance measurement.

To illustrate an approach that I feel will contribute to the needed technology of performance testing, I would like to present some information about a validation project with which I am involved. There are about 95 different jobs in the Navy to which we would like to be able to differentially assign recruits to maximize job performance.

Neither the Navy nor any of the other Services can afford to develop and administer hands-on performance tests for all jobs as we routinely validate and update our personnel classification and assignment procedures. Therefore, for a limited number of jobs, a Joint-Service working group decided to use hands-on tests as "benchmark" measures against which some less expensive, more easily administered substitute measures will be compared in large field experiments.

In the Navy, we are looking at two types of substitute measures. One involves simulation; either a paper-and-pencil test or a computer-based simulation. The second substitute measure is a set of behaviorally anchored rating scales. We plan experiments to look at the relationships among the different types of measures and between these measures and the major selection instrument used, the Armed Services Vocational Aptitude Battery or ASVAB.

The other issue that I want to raise has to do with defining what it is that we are measuring.

In the case of the Joint-Service working group that I just mentioned, we found it easy to talk about relating the ASVAB to job performance, but found it difficult to get agreement on what constitutes a comprehensive measure of job performance. When we discussed what kinds of measures we were going to use, there was concern that we were really not measuring job performance, or that we were measuring too much of it or not enough of it.

Some industrial/organizational psychologists suggest, for example, that we must look not only at whether an individual _can_ do the job, but _does_ he or she do the job? Are the individuals really there on the job? Are they absent, do they use drugs, are they motivated? That is, do they engage in "down-time" behaviors? What about the observance of safety precautions? Do they engage in "hazardous" behaviors?

All of these things make up general performance on the job. But there continued to be confusion in the working group about what aspect or aspects

40

of Job performance we should be measuring. The adequacy of any performance measure will depend on the purpose of the measurement, the measurement method used, and what is meant by Job performance. We finally decided to take a narrow focus and look only at technical proficiency. That is, we are going to test only the "can do" aspect of Job performance using tasks of a technical nature done on the job.

To put this decision in perspective, Kevin Murphy from Colorado State University and several of us at the Navy Personnel Research and Development Center tried to lay out what we meant when we talked about Job performance and we came up with a set of performance levels. These levels are most easily thought of as a type of flow chart, because it follows an individual from the time that he or she enters the Service, to the point of post-training Job performance. This is shown as five levels:

    I.   Effectiveness in a position
    II.   Behavioral components of effectiveness
    III. Interpersonal relations
    IV. Job or task proficiency
    V.   Skill and knowledge components of proficiency

The flow is from underlying skills and knowledge at Level V to effectiveness in a position at Level I. Training occurs at Level V, but practice and experience in the job environment is necessary to achieve task proficiency at level IV. Task proficiency and the other behaviors represented at Levels III and II, such as the down-time and hazardous behaviors mentioned earlier and interpersonal skills, contribute to effectiveness in a position. Effectiveness determines the value of the individual to the organization.

I believe that this concept of Job performance has application to the assessment of transfer of training. Ultimately we want to bring trainees into the job and have them be effective. That is, we want training to transfer to the job. However, we cannot determine how much transfer of training occurs, if any, unless we can specify the Job performance universe and relate performance measures to it.

It is not until we identify the concept of achievement in different performance areas, such as those represented in the levels of performance I have been talking about, that we will know what it is that we ought to measure to assess transfer of training. The main advantage of having a well defined concept of Job performance is that it will allow us to decide what the transfer tasks ought to be and how to set up the situations to measure transfer of training.

Another advantage of laying out a concept of Job performance using different performance levels is that we can see that we have to measure at different levels for different purposes.

We made a conscious decision in the Joint-Service project I described to look at task proficiency because performance at that level is most likely related to the kinds of abilities that we are measuring in the current ASVAB or in some of the new ability predictor tests that we are also investigating.

Likewise, we may have to make decisions to measure at different performance levels to: (1) diagnose learning difficulties, document achievement, and ensure transfer of training within a course of instruction, (2) measure transfer of training to proficiency in a specific job or task, (3) ensure that transfer of training contributes to global job behaviors or effectiveness in a specific position, or (4) provide quality control information or feedback to the training program.

The next item I want to talk about concerns providing more detail at the different levels of performance. An overall concept like job performance has to contain detailed performance constructs. Analytic tools are needed to specify the cognitive processes and memory structures that make up these performance constructs.

If we can analyze job performance using a cognitive task analysis procedure, we can determine what should be taught to individuals to enable them to become proficient. Performance constructs can help point out discrepancies between what is being trained and what needs to be trained, and even give us ideas on the appropriate sequencing of training. Performance constructs also can provide guidance on what to test and how to test it to meet various training measurement purposes.

Although it may be possible to achieve job and task proficiency solely through "on-the-job" training, it is usually impractical. This means we need a mix of classroom and on-the-job training. Specifying a performance construct and taking into account the resources available in the training and job environments, may allow us to make some decisions on the best mix of training environments.

What I am really proposing is that we take a "systems approach" to problems common to both validation and transfer of training work — for that we have to apply the personnel psychologist's notion of construct validation when defining and measuring job performance, and we have to capitalize on the training specialist's capability to analyze the underlying skills and abilities that are related to job performance.

In summary, we need: (1) studies of performance test fidelity that use different measurement methods and tie these methods to different performance levels, and (2) a cognitive task analysis procedure that can serve both selection and training research.

There already have been calls for a construct validation approach to criterion-referenced testing and scholastic achievement. According to such an approach, defining the performance universe and then specifying a performance construct, that is, the performance domain, is an important step in measuring transfer of training to the job. Low-cost, high-fidelity measures of performance are the key to success in this enterprise, and a cognitive task analysis procedure can help us build fidelity into our measures.

A cognitive task analysis procedure also can provide the thread that ties personnel selection to job performance via training. Personnel selection influences training by matching predictors to the method of instruction and/or course content; it influences job performance by matching predictors to the job. The same underlying cognitive processes and

42

structures that make up the job performance construct should be represented in the predictors, training content, and the job tasks.

The approach and procedures that I have advocated today imply a general need for a theory of human performance that will encompass both transfer of training and performance prediction.

THIS PAGE INTENTIONALLY LEFT BLANK

Walter Schneider, Ph.D.

Learning Research and Development Center
University of Pittsburgh

I want to talk about building what I will refer to as automatic
processing component skills. I will begin with some theory, and then apply
it to transfer of training. I will describe the distinction between
automatic and controlled processing. The interpretation of these
differences is useful for understanding how skill is acquired and help
clarify some misperceptions of the role of training, and how training
translates into performance.

I'll also talk about some empirical changes with practice, in
particular the increases in processing speed and reduced workload that are
necessary for high performance skills. Then I'll talk about the
characteristic of increased reliability. You may, in a training context,
want to train somebody not only so they can perform the task, but perform it
when they didn't have enough sleep the night before, or have a problem with
alcohol, or haven't refreshed the training, or haven't experienced the
training in six months of the task. I will also talk about the
characteristic of negative transfer, when training can influence processing
so you don't do something, or you do something that you didn't intend to do.

Finally I will go into some guidelines for developing automatic
processes, and talk a little bit about the challenge of skill training.
One characteristic of training that evolves from my work is, there's no such
thing as a free lunch. If you want high performance you are going to have to
spend a substantial amount of time in training, and developing the tasks.
I'll give you a brief description of training guidelines from this
perspective. I'll take some examples of two training systems that we're
working with at the moment, one for training electronic troubleshooting, and
the other for air intercept control.

I view human performance as being the result of two qualitatively
different processes. The first is controlled processing, which is typical
of novel situations. It is illustrated by being slow, serial, effortful ––
a conscious processing of very limited capacity and under direct subject
control. Most declarative knowledge processing involves controlled
processing.

The second mode of processing, which is more typical of the highly
skilled practitioner with years of training on the job, is automatic
processing. This is typical of situations where you do the same thing time
after time after time. It's characterized as fast, parallel, fairly

effortless, and not limited by short-term memory comparisons. I will claim that a major part of training, at least in the area of skill development, is the building up of these automatic performance skills.

To give you an appreciation of the magnitude of these effects, I will give you some examples from a basic research paradigm. The differences between these types of processing are or can be two orders of magnitude. A frequently used paradigm is a search task. In such a task you might "Push button 1 if you see the letter X or L," then you would see a set of stimuli, and you would make a response. You see a series of presentations of single letters, where you have to make a simple response to a small set of letters.

An important distinction in this work between the automatic and controlled processing is the type of mapping of the stimuli and the task. If it's a consistent mapping, the response is the same every time you perform the task (e.g., dialing your home phone number). In a varied mapping, the stimulus and response change from trial to trial (e.g., dialing random phone numbers).

In a consistent search task, you would be told in the trial "Push a button if you see X or L." L occurs and you push button 1; if a K occurs, you push a 2. Later, you see D or Q. D occurs, and you push button 1. The critical thing is, you never make a different response to the particular stimulus. In a varied search task you might search for a P or L on one trial and push button 1 for an L on trial 1. However, if on the next trial, while searching for an X or G, an L occurred, you would respond 2. The response to the L would be varied (i.e., first 1 then 2) across trials. In this varied condition only controlled processing is expected.

In the normal work environment most tasks are consistent but varied tasks occur and do cause troubles. The pressing of keys in typing, starting of an aircraft engine, or driving a car on an empty road are consistent tasks. Human computer interactions provide many examples of varied tasks. For example, different programmers use different keys to leave a program (e.g., press Esc, a function key, type "bye" or "quit"). The internal stimulus of "exit the program" is mapped to different responses. Such designs result in frequent errors and increased workload.

In the design of complex systems, there's a tendency for engineers not to worry about the consistency of the commands and the motor responses across all the methods of operation of that system. This is a varied mapping situation. The sequence that lets me do one thing when I'm using the editor will disconnect me if I'm using the modem. I have to keep in mind the goal, because the correct response is not consistent.

Let me begin with a basic effect. We've looked at these two types of processing — what the effect is of training in the controlled mode where you have this varied mapping, versus the consistent mapping where you develop automatic processing.

Using such letter search tasks, we have found no improvement in detection accuracy in the varied mapping, while finding quite substantial improvement in performance in the consistent mapping. The take-home message is that practice improves primarily on consistent tasks (Schneider, 1985, Schneider & Fisk, 1982a).

46

The statement that practice makes perfect is false. If you are practicing one of these varied mapping tasks, which we have done for up to six months, performance doesn't change much after the first half-hour. Extended training time in dealing with those varied components doesn't result in improvement, and is a waste in terms of your processing improvement.

The second characteristic that we typically want of a skilled practitioner is rapid processing of a task. I want to give you an example of a search task, and then I'll talk about the data. The important characteristic is, automatic processes typically become much more rapid. Let me give you a flavor of what happens in a trial.

On a trial, a subject searches for words from one to four categories (e.g., FURNITURE, CITY, TREE, JOB). The subject sees a display containing two words (e.g., YELLOW, ELM) and responds if there is a match (e.g., TREE to ELM).

What we find in such tasks is performance speed is a linearly-increasing function of the number of memory comparisons in the varied mapping condition (e.g., searching for one category requires .5 seconds, two categories, .7, three categories, .9). It takes about 200 milliseconds to make a semantic comparison, whether this is on day 1 or on day 25 of the task (Fisk & Schneider, 1983). However, if you are always looking for the same set of elements, and responding to them the same way, reaction time doesn't increase (i.e., search time for one to four categories is .5 seconds).

Figure 1 shows the difference in memory comparison time between controlled and automatic processing. Comparing the slopes using the negative trial responses we find that the controlled processing is nearly a hundred times slower than automatic processing of the category judgment task.

The cost of giving somebody another stimulus to look for, as long as it is responded to consistently, is negligible. In the case of the automatic mode, the marginal increase was 2 milliseconds (Fisk & Schneider, 1983). Through consistent practice there is a shift from serial to parallel processing. You can do several things at once, and we have approximately a two-order of magnitude speed-up in processing time. In many situations, particularly problem-solving or diagnosis, you want to look through many alternatives very quickly.

I should mention, any time I'm talking about the automatic mode, I am typically looking at somebody who has had in excess of 2,000 trials of performing the task. Substantial automaticity can develop by about 200 trials. Automaticity doesn't come cheap, but it can be very beneficial in terms of the nature of the processing speed-up.

This processing speed-up issue is a general, lawful relationship in psychology, referred to as the power log of practice (see Newell & Rosenbloom, 1981) — the log of the reaction time is predicted by the log of the number of trials. This holds for cigar-rolling, playing bridge, and geometry proofs. It's a very general phenomenon. General enough that I think, if you're trying to do some evaluation of a training program, you can
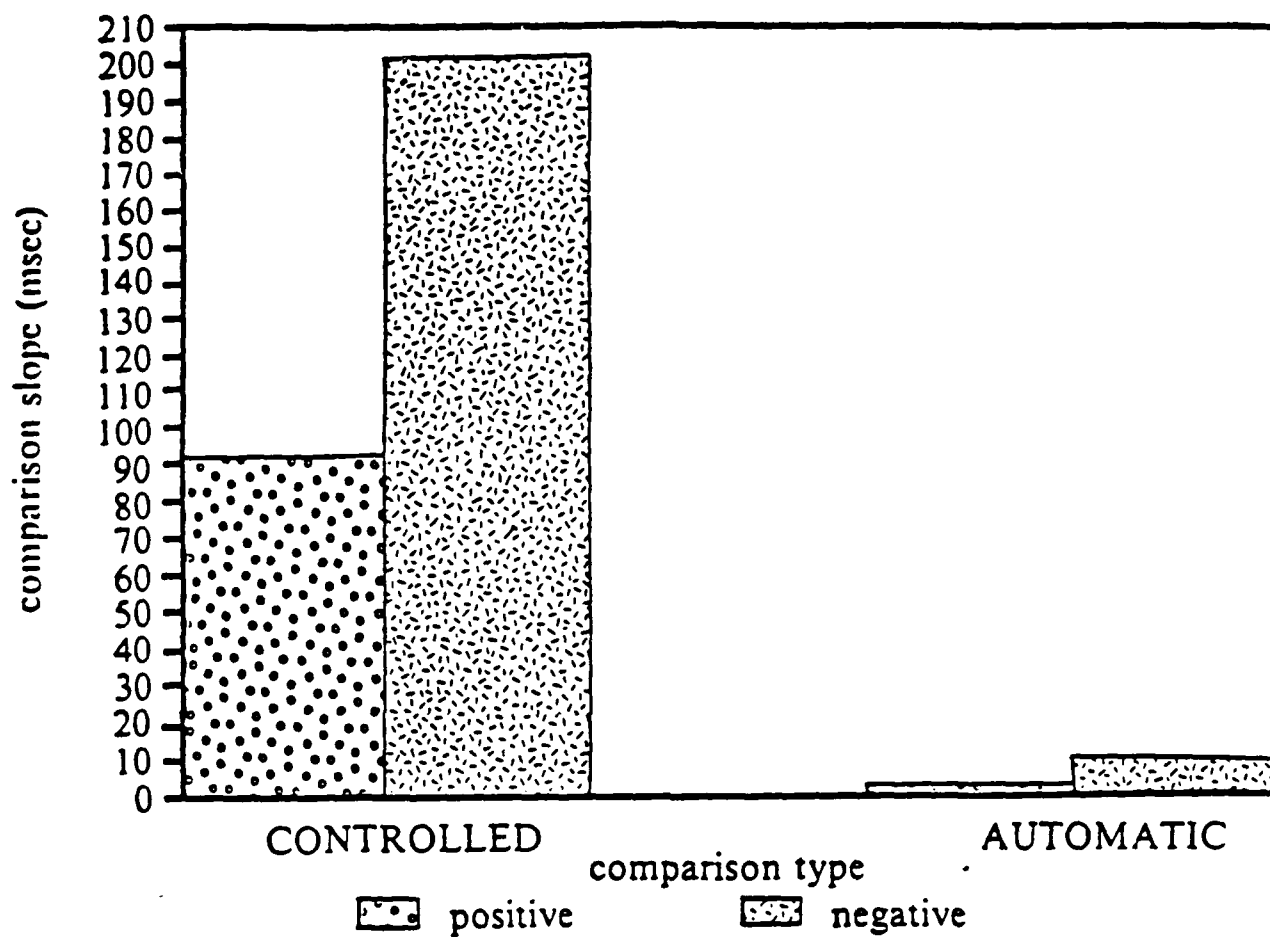
Figure 1. Representation of comparison slopes for stimuli processed using controlled processing versus automatic processing. (Adapted from Fisk & Schneider, 1983)

initially get an estimate of this slope and make a projection that will
allow you to estimate the number of trials needed to increase performance by
20 percent.

The take-home message from the standpoint of training design is that
investments in training pay off in processing speed. If the number of
practice trials is increased from 10 to 100, processing efficiency will
increase 66 percent (see Anderson in Newell & Rosenbloom, 1981). With this
kind of functional relationship, we can start building prediction into
training programs.

We have talked a little about a toolkit for a training designer. I
think these types of lawful relationships can be implemented as a subroutine
in an expert system or programming package. Then you could provide an
engineer the most effective tradeoff in training options, to give you the
best performance.

Training speedup is one aspect. An important aspect of improvements in
practice is reduced workload. This has obvious application to military
settings, where in the last 30 seconds of releasing a weapon there is a
tremendous workload. However, I believe it also has applications to much
more mundane tasks, and in particular I want to look at electronic
troubleshooting. In this controlled mode in a task using declarative
knowledge, the task is learned quickly. I want to give you a short
electronics course here. A NAND gate, which is NOT AND, has an output of
zero, if all the inputs are one, and output one the rest of the time. What
is the output if the inputs are "10", "11", "00"? It turns out, to learn
the correct outputs takes about 50 trials of training. By the end of a
training course in the classroom a student probably gets about that many
trials to practice. This is the performance level typically assessed with
paper and pencil tests.

However, the electronic troubleshooter in the field will encounter a
test point A which has a value of one, K which has a value of zero, J which
has a value of one, and T which has a value of zero. The person must
remember the test points rule for the gate.

Let's try. What's the output if the input is "KO", "A7"? What I hope
you are sensing is that this is a very different and much more difficult
task. In our experiments, it takes about 50 trials to become accurate at
performing the task, if all you're doing is filling in the answers. If you
want to do this task, and keep track of where you were in the circuit, and
not lose your temporary information, it takes about 1200 trials (Corllon &
Schneider, 1977). The normal classroom rarely provides that much practice.
With years of practice, a technician might get the training trials needed.
To make to a component skill good enough that you can build other tests on
top of that, requires this extended training. You have to, in my terms,
make that component skill automatic. Then you can integrate it and use it
in a higher level of skill. Just being able to perform the task at the low
level is of marginal utility.

An important point to note is that human working memory, where you
actually manipulate the material, is somewhere between three and four
elements — a very small number of elements. If you have to recall rules
like the operation of NAND gates while you're trying to engage in some

49

problem-solving task, you have eliminated half your memory capacity. In a
real-world setting, that elimination of memory capacity can have a very big
impact on performance, even in a situation that is not speed-stressed. In
our electronic troubleshooting classes the gates aren't that hard to learn
individually. We see a student going through a circuit very nicely until
they hit a poorly practiced NAND gate, and all of a sudden you can see their
memory deteriorate. They start randomly poking around, making measurements
in the system.

To build up a.skill usable in the world we must also worry about
workload. Being accurate is not enough. Being accurate and performing
under a heavy workload is a critical characteristic in many cases. Let me
illustrate high workload with a detection task where subjects are asked to
search for words from a semantic category. This is the task: If you see a
word from the categories mammals, birds, and fruits, tap with your left
hand, or the digit 6, 2, or 9, with your right hand. A word and four digits
are presented every .8 seconds. None of us can do this initially. Subjects
react to the task saying it is impossible. However, after eight hours of
training, you can perform digit and semantic search task as well together as
individually (see figure 2).

Let's again compare the controlled and automatic modes, using this
task. If there is a consistent response mapping, say, you're always pushing
a button to animals, that decrement in performance for the word search drops
off to the point of nearly zero with the digit task added (figure 2 right
side). Therefore, you can add another task on top of the initial task.
However, with a varied mapping the decrement in performance of doing the
word search task, given you also do the digits, is roughly 60 percent
(figure 2 left side). You can practice this for as long as you want.
Performance will not improve. In fact, it gets slightly worse.

How long does the consistent mapping task take to learn? Oh, about
2000 trials, eight hours of training in terms of doing the task, so it's not
cheap. There's no free lunch in this business; but you can substantially
improve performance. If you want somebody to operate reliably in a
high-workload situation, that extended amount of training is going to be
necessary.

We have found that the first time we introduced almost any new task,
the task interferes with any other novel task. But after a little practice,
(e.g., 50 trials), subjects start doing reasonably on the other tasks.
Processing resources become available, that is, some other tasks can be done
at the same time with little deficit. We are trying to make lawful
relationships between practice and resource load. We hope to provide the
training designer algorithms to specify when a task is sufficiently
automatic to be performed reliably in a high workload environment.

The conclusions are, initial acquisition of almost any task is
extremely resource-demanding, but with continued training, the resource
demand reduces, allowing other tasks to be combined without deficit.

An important theme here is that it's not appropriate to ask, "is part
training better than whole training?" That's the wrong way to ask the
question. One potential question is whether there is some characteristic of
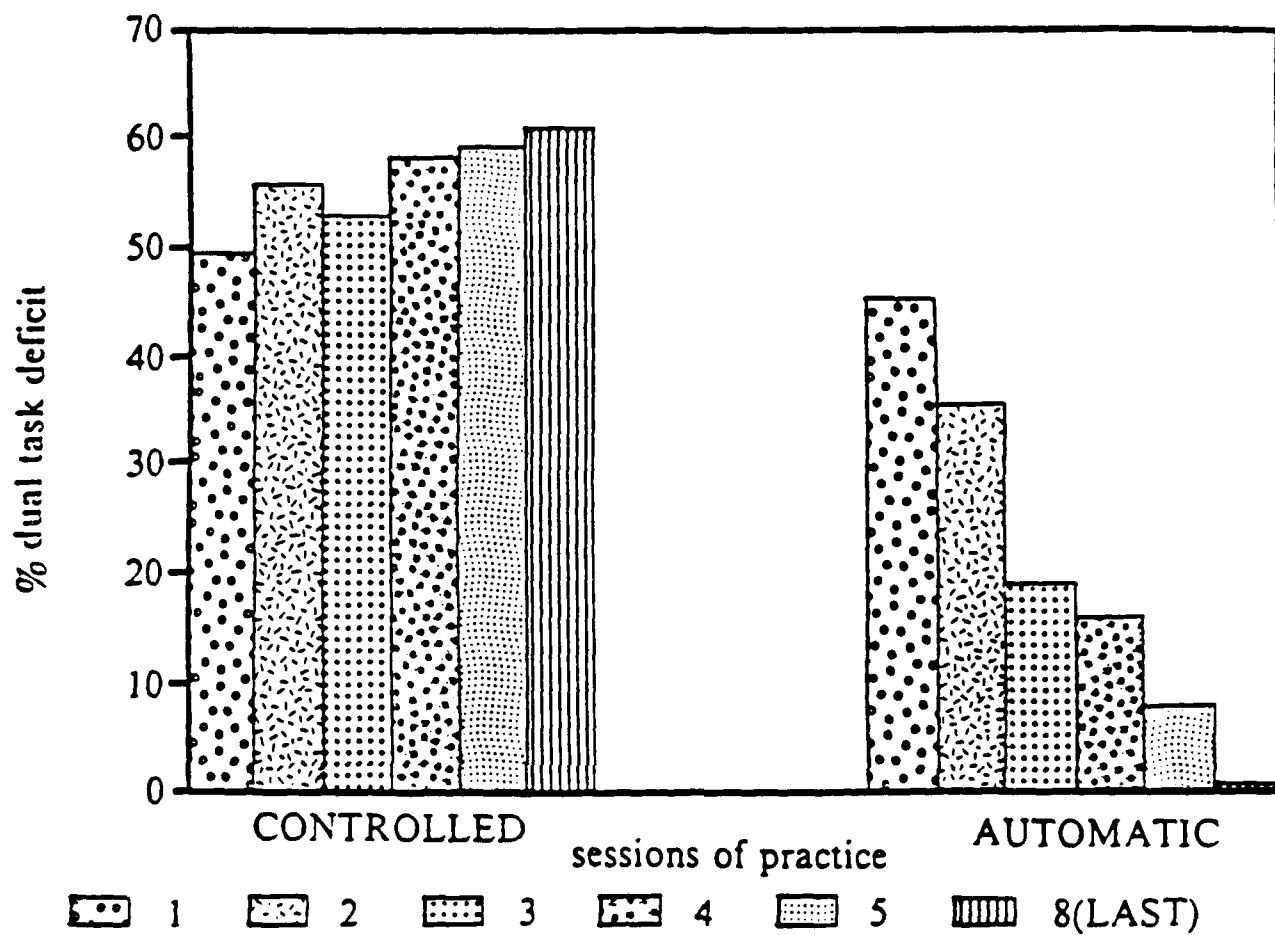the resources needed to perform the task, that can also decide when to

Figure 2. Reduction in dual task decrement across practice sessions as a function of dual task controlled processing versus dual task automatic processing. (Adapted from Schneider & Fisk, 1984)

promote from part training to whole training? If you never give the person
a way to organize the task, they can train for a very long time, and not get
any better at it.

Some military training programs like to put trainees in a high-workload
situation, and always keep them at maximum workload. We studied a version
of this using a dual task. We gave one group of people a ship control task
and a one-digit search task. The digit task was given priority and which
they were never to miss a digit. That was a modest workload, that you
couldn't trade off. Another group doing the same ship control task had a
three-digit search task, and we said, "We don't care about the digit task.
Do as much as you can on it."

The subjects under the modest workload, although they tried hard, never
got any better at the task. Doing the dual tasks was taxing and they
couldn't learn the individual tasks. In contrast, the subjects who were
allowed to trade off learned both tasks. The first let the digit task go
(i.e., missed the digits) and learned to control the ship. Then they
attended to the digit task and were able to perform both better than the
subjects who had to maintain the digit task.

The implication here is, if you put in all the parts of the task too
early, you can lose a student. That is, you give them two tasks to do at
once, while they really only have the capacity to do one, and they won't let
one of them go. They can continue to practice, and will show little benefit
or improvement. I think this might be particularly important when you're
dealing with people who have differential abilities. For example, if you
throw somebody who doesn't know English well (still dealing with
understanding individual words) into a class, and expect them to operate at
the same pace as native speakers, they may work just as hard, but can't show
a normal acquisition. They look less intelligent, although that isn't a
difference.

Let me turn now to processing reliability. We want people who can
perform well, perhaps when they haven't slept for 10 hours, or 20 hours, or
they haven't recently trained in the task (Fisk & Schneider, 1982).

We looked at the distinction between controlled and automatic processes
as a function of alcohol level. Subjects were given sufficient alcohol to
be legally drunk (.1% blood alchohol level). In the varied task, we had a
28 percent decrement in performance; in the automatic task, we had no
decrement. As a matter of fact, sometimes we got a little improvement.

The comment of, "You've drunk a lot, and you can drive home
successfully" — I won't say safely, but successfully — probably has a fair
amount of truth to it. If nothing unusual happens, if you're consistent,
and you use the normal path, you can be quite reliable, in terms of the
processing. If anything unusual happens, then you're in deep trouble.

Similar results have have been obtained for heat stress and sleep
deprivation (see Hancock, 1984). Controlled processing is quite susceptible
to these stressors, and automatic processing is much less susceptible.

There are some problems associated with automatic processing. One is
negative transfer. For example, if the location of the control that lowers

52

the landing gear is moved after years of training, pilots may experience negative transfer. In a stress situation, they drop back to the old procedure.

In the laboratory context, we examined negative transfer with a search task. We said, "Always push the button when you see A, B, C, or D," and then we switched them over and said, "Now, always push the button when you see X, Y, or Z, but ignore A, B, C, and D." In the automatic mode, it took three times longer to break the skill than it took to develop it (Shiffrin & Schneider, 1977). I might add, it took thousands of trials to develop it. In a controlled task, it turns out it took roughly the same amount of time both to develop it and to break it. This is difficult to evaluate because of errors so close to zero. There's a risk with the automatic mode that responses may occur even when they're not intended. Figure 3 illustrates the difference in initial and reversal training for controlled and automatic processing. Controlled processing required very few trials for acquisition and reversal. In contrast, automatic processing required many more trials for acquisition. However, more importantly, reversal training required nearly three times as many trials as initial acquisition.

From these results, I conclude that training programs must be designed carefully focusing on high performance skills, particularly, those that occur in high-workload situations. Suppose the task is one where the demands could change and have the potential of negative transfer. In such environments, one should redesign the task if possible. Note that negative transfer effects are most severe in very high workload situations. They are not that severe in low workload situations. To avoid negative transfer in situations that demand change, an upper-level goal state or context cue could be built in. In one context, do this, and in another, do that.

Another problem characteristic associated with automatic processing is loss of control. Apparently subway motormen hardly ever make an error dealing with the signals, such as stopping at the station (assuming they are awake). However, if they are asked to "Skip the next station," they invariably stop (Haber & Haber, 1985). If you've tried to stop at a store while driving home and gotten distracted, you know this characteristic of automatic processing. You've lost control. That loss of control is necessary to enable low-resource performance. So it is good. But there are certain situations where it can be counterproductive.

The third characteristic of problem areas for automatic processing is memory modification. How much do you learn when you execute an automatic process, versus a control process?

We have looked at the frequency with which words can be detected with automatic or control processing (Fisk & Schneider, 1984). We had subjects perform a semantic search task (i.e., look for vehicle words) and present words one to twenty times. If subjects performed a controlled search task, subjects reported that a word presented twenty times had an occurrence of eighteen times. In contrast, subjects doing automatic search had a reported estimate of zero occurrences of a word presented twenty times. Note, the subjects were semantically processing the words sufficiently to detect novel vehicle words.
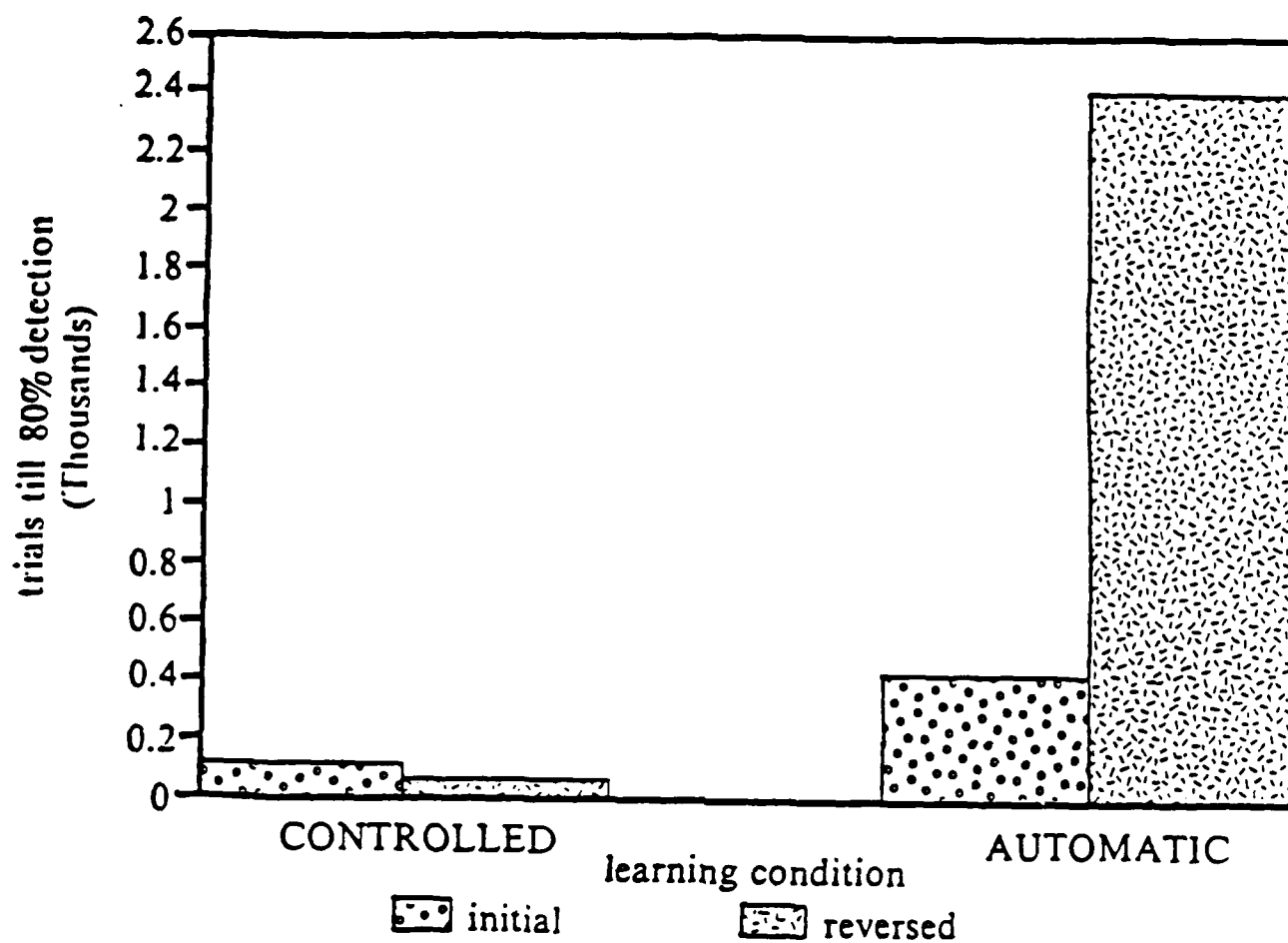
53

Figure 3. Time to acquire 80 percent detection accuracy during initial acquisition phase and following reversal of roles of the target and distractor sets. (Adapted from Shiffrin & Schneider, 1977)

These results have important implications for the training context. If you allow people to learn by rote rehearsal of the material, they may need very little memory modification in terms of the process. The ability of the person to actually remember what they did, versus what they should have done, may be greatly reduced.

To just summarize these effects, I've talked about the large qualitative differences (See Fisk, Schneider & Logan, 1987 for details). In varied mapping, you see little improvement with practice. In consistent mapping, you see very substantial improvement. So that the place to put a substantial amount of your training time, is in these consistent components (See Schneider, 1985 for review).

I've talked about the processing speed-up. We've had a hundredfold increase in processing speed. I've talked about reduced workload, and we can make a task so that it has almost no load. This is important both in the contexts of high-workload tasks, or tasks in which you're going to build another skill on top of it, like electronic troubleshooting. I've talked about processing reliability. We want this task to go on, even if the person isn't in the best of shape. I've talked about the negative effects, that we can have a greater problem of interference effects. We can have some loss of control, and less memory modification, for these automatic processes.

I have a model which assumes that processing goes through four phases (Schneider, 1985). I think most classroom training gets to the second stage, still basically in the control processing mode. In many tasks, reaching the third stage, where you have to at least generally attend to the task, can be important. To do that typically requires a substantial amount of training. The third stage begins after about 50 trials of doing the task. For example, for many of the component tasks performed by air traffic controllers or air intercept controllers, they haven't had 50 trials of doing that task even after a six-week training program.

There is a crisis and a challenge in skill training. It's a challenge, because we know that with the right kind of practice we produce dramatic performance improvements. We can do it in the laboratory setting, and in certain applied settings.

It's a challenge because it takes a lot of time. After ten years of experience on the job, you can see these performance changes. Compare the expert and the novice air traffic controller. The expert sits there quietly making commands without looking overloaded. The novice is sweating profusely, can't figure out what is going to happen next, and is overloaded. How do we compress experience? The military has this problem, and it doesn't have the luxury of waiting ten years to grow a new controller.

In skill acquisition, from my perspective, automatic component skills must be developed for those tasks that keep reappearing. A critical aspect of training is building up those components, and developing the time-sharing skills that allocate control processing to the inconsistent or poorly learned task. For example, in the case of electronic troubleshooting, keeping track of the voltage levels throughout the circuit. People need some experience in that, but it doesn't help to give them a lot of experience.

55

One of the processes that we are using to facilitate automatic control
is this time-compressed training. We provide the learner with many
consistent trials performing the task. In the case of air traffic control,
the critical aspect for air intercept control is bringing two airplanes
together. The critical task is projecting the trajectories of those
airplanes and getting them to the right point. If you intercept in real
time in the simulator, a single intercept takes about 20 minutes. In this
case, we can compress time by a factor of 100:1. Experience with 2,000
maneuvers in each of the particular areas of air intercept control could
take five or ten years on the job. We can provide 2,000 trials on each of
the critical components in compressed time in about a week. We practice the
components and then deal with the total task. This does not suggest that
normal procedure is eliminated. Rather, initial training is performed at a
time compression that is optimal for learning, not necessarily the time
scale of the real world.

It is difficult to do 2,000 trials of anything, and keep going.
Therefore motivation is a very important part of training. In my own case,
the most cost-effective piece of equipment in my lab is a $15 noise
synthesis chip that makes interesting sound effects. It keeps subjects
going after thousands of trials. In the electronic troubleshooting task, we
use competition to keep students working for weeks.

I'll just briefly present some of the guidelines for designing training
programs (See Schneider 1985, Regian & Schneider, 1987). The first one is
"Present information to promote consistent processing." There are a number
ways of doing that. One is that you can use a verbal rule. In the case of
the electronic gates, it's on if any input is on. You can use an analogy.
You can provide examples. You can provide a table. The function of all of
this is to produce a consistent execution during the first phase of
training. If the student doesn't see the consistency of how to deal with
that task, you are not going to see much of an improvement in practice. In
many complex tasks, for example in air traffic control, it is not easy to
see those consistencies. You have to design the tasks so that the
consistencies are apparent.

Next, "Introduce novel components singly, and do not overload
processing during initial acquisition." Now, remember that if initial
resources do not exist, extended practice will lead to little improvement in
performance. With the available resources, you see a reasonable performance
improvement function.

The third point is, "Design the task to allow many trials of the
critical skill." In the electronic troubleshooting, for example, we give
them a lot of trials in dealing with gates, in terms of a simple prediction.
The goal is to decrease response time as a function of training. More
importantly, many real world tasks must also be accomplished under high
workload. In the case of electronic troubleshooting, that workload is
imposed by keeping track of other elements of the task.

Four, "Maintain motivation throughout training." With training using
hundreds of trials, motivation is very important. With subjects in the lab

we find that if we don't worry about the motivation, they will show a nice improvement function for a while, but after about the fourth or fifth hour, learning levels off. Real training programs generally all go much beyond four or five hours.

Last, "As automaticity develops, we utilize the available resources to integrate and enhance component tasks." So that, as the resources become available, training can use those available resources. In the electronic troubleshooting, training moves from gates alone to training in the circuit. In some sense, training time is wasted if you spend too much time with components.

Human performance changes drastically with practice. I believe behavior is the interaction between two quantitatively different forms of processing, referred to as automatic and controlled. This distinction has implications for training. In particular it suggests that we focus practice on the consistent components of the task. We design the training environment to allow many training trials, and we promote to higher levels of skill training on the basis of workload, in terms of processing.

References

Carlson, R.A. & Schneider, W. Learning and use of causal relations: Representation, direction of judgment, and reversibility. Unpublished publication.

Fisk, A.D., & Schneider, W. (1982). Type of task practice and time-sharing activities predict deficits due to alcohol ingestion. _Proceedings of the 26th Annual Meeting of the Human Factors Society. 26_, 926-930.

Fisk, A.D., & Schneider, W. (1983). Category and work search: Generalizing search principles to complex processing. _Journal of Experimental Psychology: Learning, Memory, and Cognition. 9_, 177-195.

Haber, R.N., & Haber, L.R. (1985). Case Study: Automatic and controlled attention; subway motormen and controllers. Presented at the 26th Annual Meeting of the Psychonomic Society, Boston, MA.

Hancock, P.A. (1984). Environmental stressors. In J.S. Warm (Ed.), _Sustained attention in human performance_, New York: Wiley.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), _Cognitive skills and their acquisition_. Hillsdale, NJ: Erlbaum.

Regian, W. & Schneider, W. (1987). Assessment procedures for predicting and optimizing skill acquisition in extended training. In N. Frederiksen (Ed.), _Diagnostic monitoring of skill and knowledge acquisition_. Hillsdale, NJ: Erlbaum.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. _Human Factors. 27_, 285-300.

Schneider, W. (1985). Toward a model of attention and the development ofautomaticity. In M. Posner & O.S. Marin (Eds.), _Attention and performance XI_ (pp. 475-492). Hillsdale, NJ: Erlbaum.

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.

Paul Thayer, Ph.D.

Department of Psychology
North Carolina State University

I'm going to talk about a very different transfer problem, the climate within an organization which may inhibit or facilitate transfer of training. But, first, a little background.

Since going to my new life as a department head at North Carolina State, I have had a number of students working on interpersonal skills for managers — teaching how to motivate, how to set goals, how to discipline, all these different kinds of things.

We use a behavior-modeling training program, based rather loosely on Bandura's social learning theory, concerned with teaching interpersonal skills. For a variety of reasons, this kind of training is less likely to transfer than the typical kind of skill training. There may be factors within the organization which prevent transfer: expectations of subordinates as to how the trainee should behave, production demands, etc. I want to focus on these kinds of problems, instead of the traditional transfer problems.

In interpersonal skills training, as in all training, it's very clear that you need to do needs assessments. Unfortunately, in training in the life insurance industry (where I worked for 21 years), in almost any industry, and most likely in the military, there's an awful lot of training that gets done without any needs assessment whatsoever. This, of course, results in faddism. Quite frankly, I think the Army's move into organizational effectiveness was in part motivated by some faddism as opposed to a basic and thorough needs assessment. As Bob Mager has pointed out so well, many performance deficiencies are not training problems. They are selection problems, or motivational problems, or some other kind of problems. We do any awful lot of training when we should be doing something else.

As you know, needs assessment should tell us what the goals and objectives of our training might be, the criteria to be used in the evaluation, and the content and the methods of training. Bill McGehee and I have developed a basic trilogy for needs assessment: organizational analysis, task analysis, and person analysis. Because task analysis and person analysis are so familiar, I want to spend more time talking about organization analysis.

In organization analysis, the organization's objectives, resources, and the climate for training or transfer are carefully analyzed. The organizational creed, or goal statement is examined. Organizational resources in terms of humans and equipment are examined. The acceptance of training within the organization is evaluated, and so forth. Unfortunately, getting all this information and putting it together is an art, not a science. Thus, our training may fall short and fail to transfer.

When we find out there's a need, we train, we test, we find out what people know what we taught them in the training. We ask them to behave in the way we taught them to behave in the training. But often, when we look at them on the job, nothing happened. We wonder what's wrong. What prevented transfer? Sometimes we find the behavior did transfer, but it had no effect. What happened? Were there organizational factors that prevented transfer?

Let me give you an example. In 20 years, 30,000 foreign engineers were trained in the United States, with the objective to train them here and have them return to their countries of origin. Well, those engineers are still here, competing with American engineers. An appropriate organization analysis could have shown the foreign engineers were being trained to work in an environment that doesn't exist in their own country.

Another example: Police training typically emphasizes the use of firearms, filling out reports, and legal requirements. The task analysis shows this, but an organization analysis would point out organizational expectations about interpersonal relationships with the public, positive image, et cetera. If you don't include these in your training, the training will fail.

Or, look at the training of the hard-core unemployed which occurred many years ago. They were taught how to perform on a job, not about other expectations, such as being punctual, working every day, respecting the foreman, and so forth. If these skills aren't taught, the job skills won't have a chance to transfer.

Let me discuss one particular aspect of all of this: the climate for transfer. I will be borrowing heavily from Irv Goldstein in what follows.

Basically, we want to learn about organizational facilitators and inhibitors of transfer. With such information, training can be modified to innoculate employees against potential problems.

We know how to go out and get knowledge and skills and attitudes required on the job. We're not very good at providing inputs to instructional designers concerning the transfer climate.

To overcome this deficiency, Irv Goldstein is attempting to develop a method for systematically assessing transfer climate. The same subject matter experts used for developing task analyses will be interviewed to find out about facilitators and inhibitors for transfer.

He is using the Katz and Kahn open-system model. Any organization, according to that model, is an open system which has five major dimensions: support, maintenance, production, adaptive, and managerial dimensions. The

support system is important in providing support for the resources of the organization. Maintenance are those various ways of attaching the people to work, keeping the system going. Production is the actual display of performance. The adaptive system is the long-term provision within the organization for adapting to change as times and conditions change in the outside world. The managerial, of course, is the whole business of supervising policy coordination and control.

For example, under maintenance, you might ask, is training made available when you need it? Do employees leave because developmental training is not available? Under production, do job incumbents get a chance to practice learned behavior in the organization, or does it only come about in the event of an emergency? Do employees in our department use safety behaviors taught to them in training when it interferes with a production standard?

If we can accurately assess the climate, what do we do from that point? If we find we have a negative climate for transfer, what kind of strategies or options do we have at that point?

There are a number of very definite approaches to take with management. Trained behaviors must be reinforced on the job. A trainee who comes back into an unappreciative environment for skills trained will be ineffective. Strategies must be taught which will enable the trainee to transfer skills despite that environment. Another problem occurs when the skills taught are not used for several months. What can we teach the trainee to keep the skills "alive?"

I think, if you have an accurate assessment of the transfer climate, you may be able to, in essence, innoculate the trainee — or provide support systems for the trainees so that even if the organization itself doesn't change, the trainees can go there with the appropriate mechanisms to make sure that the behavior does transfer.

Irv's work, which I plan to try out, is still in the early stages of development, but I believe it is the first systematic breakthrough in an important area. I think it will help give us better understanding and a means of anticipating transfer problems more systematically. That will provide us with the means to avoid or overcome inhibitors, and to capitalize on facilitators.

THIS PAGE INTENTIONALLY LEFT BLANK

Robert Swezey, Ph.D.

Behavioral Sciences Research Center
Science Applications, Inc.

I would like to comment on five specific areas within the general
context of transfer of training research. Some of these areas are very
broad, others are more specific.

1. Let me begin by saying that, in my opinion, the term "training" is
so broad as to severely confuse many of the discussions that are devoted to
it. Training in art appreciation is very different from training in
electronic troubleshooting. Training in basketball foul-shooting is very
different from training in cognitive psychology. Training in woodcarving is
very different from training in 18th century French literature. In my
opinion, preferred approaches to training, and therefore to transfer, are
inexorably confounded with content. This is, of course, not new. A lot of
people have made this point previously. We have discussed it here at some
length.

A major problem in the area of training and education that has been
discussed literally for centuries, involves the relative importance of
pedagogy versus content. In my opinion, this issue is far from being
resolved, and continues to haunt our investigations of transfer of training,
both in the military and in civilian environments. I personally feel that,
in general, we do not adequately provide for mastery of the necessary
content material by instructional developers prior to and during the
development of training programs. In many military training environments
with which I am familiar, instructional developers are not themselves
subject matter experts. Further, in my opinion, they do not adequately
consult subject matter experts in their development activity. I feel that
instructional developers should be subject matter experts who are trained in
pedagogy or education (or who bring instructional experts in from the
outside), not vice versa. In my opinion, this is a big mistake.

Of course, it's possible for a smart person to learn either (or both)
of these disciplines. However, it could take years for a training expert to
master a unique content domain, such as international mortgage banking, for
example. So, where does this leave those of us who are in the training
research community? In my opinion, we need to develop interactive design
taxonomies, which combine such approaches as the Department of Labor's
occupational codes, with the types of tasks and skills involved, and overlay
this with appropriate training strategies for each unique combination.
Until we start working toward something like this, we are going to be
continually criticized for not having our act together, and, in my opinion,

63

criticized with some justification. Research has been conducted for decades in this area, and a lot of wheel-spinning and reinvention is going on. We need to codify this vast training literature in some defensible way, so that we stop reinventing the wheel, and can generalize, and so that we don't have to conduct unnecessary transfer of training studies, in areas that we already know about.

2. Secondly, we need to arrange the situation so that hardware and software, in the training development world, does not drive training, but vice versa. This is, in my opinion, a big problem area. The potential power of the new technology in training and instruction (such as teaching machines, videodiscs, simulators, CAI, so-called "intelligent" CAI, artificial intelligence techniques, etc.) is awesome, but we can't let it drive our instructional development efforts.

I recently ran into what I consider to be a very cogent article on this topic by Derek Bok, who is the President of Harvard University. This article was excerpted from Bok's annual report to the Harvard Board of Overseers. (I, by the way, was amazed that the President of Harvard University would devote his report to the Board of Overseers to the topic of transfer of training.) In that article, Bok cited a number of observers of the training and educational technology scene, in what is essentially a "good news/bad news" type of format. Let me extract from some of his comments, for the purpose of our discussion here.

First, the good news. According to Steven Muller, President of Johns Hopkins University, due primarily to the advent of CAI, "We are, whether fully conscious of it or not, already in an environment of higher education that represents the most drastic change since the time of the University of Paris, some eight or nine centuries ago." Similarly, according to Ray Neff, Director of Computer Sciences at Dartmouth, "Because of the speed and accuracy of the computer in performing computations and processing information, what Ph.D.'s did 25 years ago can be term projects for Dartmouth students today." And, over twenty years ago, Patrick Suppes made the following statement: "One can predict that, in a few more years, millions (of students) will have access to what Philip of Macedon's son Alexander enjoyed as a royal prerogative: The personal services of a tutor as well-informed and as responsive as Aristotle."

Bok goes on in his article, however, to temper this euphoria with a certain amount of skepticism (i.e., Now, the bad news). According to Richard Clark at the University of Southern California, "Five decades of research suggest that there are no learning benefits to be gained from employing different media in instruction, regardless of their obviously attractive features or advertised superiority. The best current evidence is that media are mere vehicles that deliver instruction, and do not influence student achievement any more than the truck that delivers groceries causes changes in our nutrition."

Bok himself states that, "Experience should make us wary of dramatic claims for the impact of the new technology. Thomas Edison was clearly wrong when he declared that the telegraph would revolutionize education. Radio, also, did not make a lasting impact on the public schools, even though foundations gave generous subsidies to bring programs into the

classroom. Television met a similar fate in spite of glowing predictions heralding its powers to improve learning."

In each instance, according to Bok, "Technology failed to live up to its early promise for three reasons: Resistance by teachers, high cost, and (most importantly) the absence of demonstrable gain in student achievement. There is as yet no clear evidence that computers and videodiscs will meet a happier fate."

According to Bok, "The educational benefits of technology remain in dispute. There is still little proof that these devices yield lasting improvements in learning. Many studies purport to find such gains. But, most of them can be explained on the grounds that students using computers were temporarily motivated, by the sheer novelty of the machines, or that more effort and better teaching went into the computerized courses than were devoted to the conventional classes with which they were compared. Thus, learning improvements that early investigators reported shrank to virtually nothing, when the same teacher taught both the experimental and conventional classes with comparable amounts of preparation. Similarly, the gains achieved in computer experiments lasting less than four weeks dropped by more than two-thirds, when the experiments continued beyond eight weeks, and the novelty of the new technology wore off. Undaunted by such obstacles, educators and high-tech companies spent huge amounts of money to prove the skeptics wrong. Control Data Corporation reputedly invested almost a billion dollars in the computerized college curriculum, PLATO."

What does all this mean? In my opinion, this means that, at some level of detail, we've got to guard against the tail, essentially, wagging the dog. I personally have some data to support this contention. In a recent series of studies comparing interactive, computer based systems of instruction, with two much less expensive methods of instructional delivery in electromechanical troubleshooting, no differences were found in mastery or transfer across instructional delivery methods when the time spent in training was controlled, and the material was identical. This occurred for two tasks, of varying complexity, both immediately following training and after a one-week retention period.

The point here is that, in my opinion, our research should deemphasize hardware, and should instead, emphasize the specific training requirements for learning and optimal instructional strategies.

3. Third, let me turn to a topic which I believe is dramatically underemphasized in our research on training: the use of concrete goal setting. I realize that there is widespread difference of opinion as to what transfer is. Is it the difference between performance of trained versus untrained persons? Is it the difference between obtained transfer and the maximum possible transfer? Is it a measure of learning rate? Regardless of these definitions, all have in common the perspective of how later performance is affected by current training. One common way of enhancing downstream performance which has been widely researched in the industrial (but not the transfer of training) world, is via the application of concrete goal setting. There's a body of data that testifies to the superior performance effects of concrete versus abstract goal-setting. However, there's very little work out there, (insofar as I am familiar) on this topic in the domain of transfer of training.

A singular exception is a program that was developed at the Office of Personnel Management. It is called PAPA — Participant Action Plan Approach. PAPA is a technique which requires that training participants commit in writing to concrete performance goals, which they themselves initiate, and which are based upon the training they receive. The extent to which these goals are met in the subsequent performance environment is the training evaluation criterion. I want to mention this area as one that I think may be a potential gold mine for facilitating transfer of training.

4. The fourth topic I would like to comment on involves the development and the use of models, whose purpose is to predict potential transfer of training without actually conducting transfer experiments. To my mind, this is a major area of endeavor which could produce substantial results if such models were thoroughly developed, implemented, and validated. Models of this sort have been around for a long time — Osgood's transfer surface is an example. However, they have not been implemented in practical ways in an applied environment.

Of note are the recent ARI and Naval Training Systems Center efforts on this topic. These studies have involved a series of models formerly known as TRAINVICE, and more recently known as DEFT, (e.g., Device Effectiveness Forecasting Technique). I have passed out a report in this area to participants at this meeting. Basically, there were several such models developed by a variety of developers. The report that I passed out is a commentary on four of these models.

Presently, the research and development needs of these models involve four major problem areas as follows: (1) the theoretical constructs of the models, (2) their mathematical formulations, (3) measurement issues involving the validity, reliability, and precision of the models, and (4) their convenience of application and acceptability to the user.

Regarding the theoretical construct of the models, the four TRAINVICE models reflect the following basic assumptions: they attempt to find out what people know now, to determine what they need to know, to compute the delta between these two positions, and to overcome that delta with training techniques that deliver training content designed to assure learning. Further, the models attempt to assess the extent to which the similarity between the device and the parent equipment is sufficient to permit adequate transfer to occur.

Regarding the mathematical formulation, all TRAINVICE models provide a mathematical formula intended to forecast transfer. There is, however, no discussion in the models to the possibility of negative transfer. Further, the models assume that physical and functional similarity between the device and the operational equipment are equally important in facilitating transfer.

It seems to me that there are additional areas that need to be investigated. The TRAINVICE models have never been validated. A variety of studies need to be conducted to test these models in applied settings. Further, such measurement models must be convenient to implement. The TRAINVICE models as they currently exist, are extremely inconvenient to implement. This is another area that requires major attention.

Such approaches as the TRAINVICE models are in their infancy. There's a lot of potential merit here. The recently revised DEFT models developed by the Army Research Institute are in fact, a step in this direction.

5. The fifth and final area that I'd like to comment upon, involves the topics of technical documentation and job aids.

Over 20 years ago, Wolf and Berry described a job aid as something that guides an individual in the performance of a job, so as to enable him to do something which he had not previously been able to do, without requiring him to undergo complete training for each task. Others have defined job aids as documents or devices that store information which is required to perform a particular operation or set of operations, and which makes the information available on the job.

Job aids differ in known ways from other system elements, like tools and training. Job aids differ from tools in that tools generally do not store information and make it available on the job. Job aids differ from training in that training is designed to encourage the learning of a particular skill, whereas job aids are designed to assist in the performance of that skill in the work environment.

I'll mention briefly a couple of studies in the job aids realm, done by Elliott and Joyce in the '60s and '70's that provide a fair amount of data to testify to the utility of job aids, particularly in complex environments.

One study involved a job aid that was designed for use by low-aptitude, novice technicians on a radar system. In that study, the technicians using the job aid were capable of reducing the time required to isolate and correct malfunctions by as much as 50 percent, over the time required by highly trained technicians using conventional procedures.

A second study compared high school students, using job aids, to highly trained electronics technicians, using standard manuals. In that study, high school students were given 12 hours of training on identification of electronic components, using a volt meter, and basic soldering procedures. The technicians, on the other hand, had considerable formal training in electronics, the majority having three to six years of field experience. Both groups were assigned to tasks involving complex maintenance of electronic equipment. The students using the job aids outperformed the technicians in every phase of the study. One subject in that experiment, in the skilled technician group, required 12 hours to troubleshoot a radar system, performed 133 steps, and referred to 41 different sections of 8 separate manuals. A high school student, using the job aid, required five hours to perform 35 steps, and referred to only three different manuals. The bottom line is that a fair amount of data exists to testify to the potential benefits of well-designed job aids in reducing the need for extensive training on new tasks.

Jack Folley, in 1975, addressed the issue of comparing performance based exclusively on training to that based exclusively upon job aids. He hypothesized that if we allocate the entire requirement to training, we might achieve a relatively low level of performance, because of the large amount of information required to be learned in a complex skill area. Similarly, if one allocates the requirement entirely to job aids, we also

may get a relatively low level of performance, due to the fact that the performer may become tied to the aid. If the aid contains an error, the performer gets lost. Folley has come up with a gradient that describes the tradeoffs between these two, and generates a hypothetical optimal mixture of these two functions.

The Navy's EPICS program uses an alternative approach, which also includes the use of job aids. It provides for so-called "enriched" and "hybrid" job aids. A "hybrid" aid is one that presents troubleshooting information, both in deductive and directive formats, whereas an "enriched" aid is one that presents additional job information to facilitate the transition between directive and deductive formats.

Because of job aids, the definition of transfer may need modification. In some sense, the process of transfer implies at least two phases. The first is when general information is recovered from prior experience for transfer to a new situation. The second is the direct application of that prior experience to current performance or to a current learning situation.

A job aid is often used as an aid to memory, by not requiring a person to recover something from past experience to apply to the current situation. It unburdens an individual's memory, and therefore facilitates transfer which otherwise, might not be possible. The point is, should "transfer" be defined only in terms of having to tap one's memory to retrieve information?

I suggest that improved technical documentation, including job aids, is an area where major strides can be made in DoD. Certainly, the demonstrated effectiveness of artificial intelligence-based diagnostic information systems is one major area where improvement has been made. The use of handheld tutors in the on-the-job environment is another, yet much more work remains to be accomplished in these areas. I think that improved technical documentation can reduce many of DoD's transfer of training problems.

I have commented in five areas of training and transfer that I think are important to those of us who are in the research community. I wish to emphasize that these and other similar areas of interest must be developed and tried out in applied settings before any meaningful conclusions can be drawn.

SUMMARY RECOMMENDATIONS

Following the presentations, all participants summarized main points
from the workshop and made suggestions for future research and related
activities. Following are the primary recommendations raised during the
summary discussion. They are the recommendations of one or several
individuals, and do not necessarily reflect a consensus of opinion. The
recommendations have been combined by specific area.

ISSUE	RECOMMENDATION
Definitions	There is a need for standard, operational definitions of several terms, including "transfer of training," "far transfer," and "near transfer."
Interventions	For the benefit of researchers who are uninformed of past attempts at interventions, a listing should be prepared of interventions that have already succeeded (or failed) in a military setting.
	A listing of needed interventions should be developed, as well.
Measures	The research community should develop a listing of measures of transfer, including the data that are needed in order to calculate them. When measures are not currently available, the research community should determine what methods are needed to develop them. This would include measures for both actual job performance and prediction of job performance.
Research Needs	There is a need for investigation of team performance, the necessary mixes of skills on teams, and a team person/job match optimization system.
	The relationship between job performance and training conditions needs to be investigated further, and should include a substantial data collection effort.

There is a need for a theoretical base for the development of unpracticed skills.

High-fidelity, low-cost performance measures need to be developed to assess transfer of training and to validate predictor tests.

Design of Instruction

In developing intelligent job aids and other instructional devices, researchers should anticipate the transfer situation.

Systematic instructional development processes, including feedback and monitoring systems, need improvement.

The training community should make known the types of tools that are needed, and the research community should work on the algorithms for the development of tools.

Researchers should become knowledgeable about the content matter of the training task (e.g., electronics) before attempting to devise training programs.

Delivery of Training

There should be development of job aids (possibly computer-based) aimed at the changing expertise of the trainee.

The implementation of training programs should be monitored more effectively.

Training environments should be structured to allow for sufficient training trials, to promote higher levels of skill development.

Applications

Researchers should be more cognizant of the potential applications of their work. What applications do the participants anticipate from their work?

Funding

Academic-based researchers need encouragement, possibly through a funding initiative, to produce more research with direct application to transfer issues.

70

**Workplace/Equipment
Design**

The workplace environment and equipment design must be considered when developing job aids and related devices.

**Sharing Research
Results**

Psychologists should communicate research results to the user community using terminology appropriate for that group.

Basic and applied researchers should communicate more effectively with one another to stimulate work in both directions.

Task Analysis

Analysis of tasks should be based on underlying cognitive and psychomotor processes needed for performance of specific functions.

Task analysis is needed to specify the cognitive processes and memory structures in performance constructs.

Future Discussions

There should be more opportunities for scientists from the military laboratories, and both academic and non-academic settings to discuss current research and future research needs. It was also recommended that representatives of other professions (e.g., job aid development) would provide useful comments.