Learning About Equipment from Technical Documentation
A Basic Comprehensible Writing Aid
Final Report

David E. Kieras
University of Michigan

Technical Communication Program
Technical Information Design and Analysis Laboratory
2380 Bonisteel Blvd.
Ann Arbor, MI 48109-2108

Technical Report No. 31 (FR-89/ONR-31)

January 10, 1990

This research was supported by the Office of Naval Research, Cognitive Science Program, under Contract Number N00014-85-K-0138, Contract Authority Identification Number NR 667-543. Reproduction in whole or part is permitted for any purpose of the United States Government.

Approved for Public Release; Distribution Unlimited
This is the final report for a two-part project. Theoretical and empirical work was conducted on the topic of the role of a mental model or how-it-works information in learning to operate equipment. A general conclusion is that high-quality training on specific procedures is generally superior to training limited to system knowledge, which was relatively difficult to learn and to apply. Additional work was also done on learning procedures from text, which further extended production-system models as an account of procedural learning. The second part of the project was further work on a computer-based aid for comprehensible writing of technical materials. Applications of the project work and various problems encountered are summarized.
ABSTRACT

This is the final report for a two part project. Theoretical and empirical work was conducted on the topic of the role of "mental model," or how-it-works information, in learning to operate equipment. A general conclusion is that high-quality training on specific procedures is generally superior to training limited to system knowledge, which was relatively difficult to learn and to apply. Additional work was also done on learning procedures from text, which further extended production-system models as an account of procedural learning. The second part of the project was further work on a computer-based aid for comprehensible writing of technical materials. Applications of the project work and various problems encountered are summarized.

Project Goals

This research contract actually encompassed two projects. The original one, under the title Learning about Equipment from Technical Documentation, was concerned with empirical and cognitive modeling studies of how people learn to operate equipment from the kind of information contained in technical documentation, namely, information about operating procedures, and information about how the device works. The goal was to understand how knowledge about equipment could be effectively presented, both in terms of what the user is supposed to do procedurally, and what about the equipment itself should be conveyed. The second portion of the project was added later for administrative reasons; this was a separate line of work on a computerized aid for comprehensible writing, a computer program intended to provide feedback to the writers of technical documents concerning comprehensibility problems. This sub-project was sponsored by the Navy Personnel Research and Development Center.

Work Accomplished

The Role of Mental Model Information

Questions from earlier work. This project started where earlier work left off (see Kieras & Bovair, 1984) on the topic of the role of mental model information about how the device works. At that point we had demonstrated empirically that training in mental model information for a simple control panel device resulted in much improved performance in paradigms where people were explicitly taught or had to infer procedures for operating the device. A simulation model had been constructed of the mental model reasoning process, (see Kieras, 1984; Kieras, in press) and some work was done during the period of this project showing that the model accounted for some aspects of the data. This work was submitted to two different journals, with negative outcomes. Overall the reviewers did not seem to agree that the approach was worthwhile and did not consider that the comparison of the model to the data was adequately impressive. The data in question were individual inter-response latencies for elementary actions on the
device (e.g. pushing a button) where each subject where each subject contributed one trial on each problem. Since the model was a model only of correct performance, only about half of the response sequences in the data could be compared to the model. Of these, quantitative predictions made from the model and various nuisance variables could account for roughly 40% of the variance in the inter-response latencies for individual subjects, which as it happens is close to the maximum variance that could be accounted for given the reliability of the data. Thus, it would appear that perhaps the reviewers simply did not appreciate that accounting for problem-solving data in such detail was an accomplishment, or perhaps there has been a shift in paradigm where simply being able to model some data is not considered publishable anymore. But in any event, because new data were being collected, I felt that further effort to publish an admittedly limited set of results was not worthwhile compared to work using more comprehensive and reliable data sets.

So a new line of studies was begun, with the idea that the modeling of mental model processes would be done in service of a larger scale goal, rather than just demonstrating that such models were possible. A set of such larger goals appeared during work on my contribution to the ONR- and ETS-sponsored conference on “Diagnostic monitoring of skill and knowledge acquisition” (Kieras, in press). The mental models trained in the earlier studies were in fact logically incomplete, and contained highly specific rather than generic information. This differed sharply from the assumptions that I had made in the cognitive simulation model, and so it was clear that the content of the mental model training did not agree with the logically complete specification of the mental model represented by the simulation. Perhaps a cleaner set of data, and better fit of the model would be obtained under conditions where subjects were trained in the same mental model content.

Furthermore, during the same work I realized that the earlier studies had compared mental model training to conditions in which subjects either inferred the procedures on a trial and error basis, or were given training in rather sub-optimal procedures. That is, the original study in Kieras and Bovair (1984, Exp. 1) trained subjects in procedures that were highly overlapping and highly redundant, but this was not made explicit to the subjects. Our transfer work (Kieras and Bovair, 1986) was based on the realization that this overlap explained the rote training data. Thus, the procedure training that subjects received was rather sub-optimal. It turns out that a complete and accurate procedure for the control panel device can be stated in almost the same number of steps as just one of the ten procedures that subjects were trained on.

So, a better view of the Kieras and Bovair (1984) studies is that mental model training was shown to be superior to both trial and error and grossly inferior procedure training. The problem is that military training practices correspond to neither one of these radically inferior conditions, however much their need for improvement. The new studies compared mental model training to high-quality procedure training.

A final goal of the new line of studies was to get performance up to a higher level than previously so that a larger subset of the behavior sequences could be compared to the simulation models.

New control panel device studies. A series of three careful studies on the control panel device were undertaken; these studies were quite sound methodologically and produced very high-quality data. In these studies different forms of mental model training were compared to a high-quality rote procedure training condition. The results were ambiguous for understanding the value of mental model-based training. Throughout all of the studies, the rote procedure training was by far the best condition in terms of speed of learning and quality of performance. The different mental model conditions proved to be indistinguishable, except in the earliest trials, generally poor in overall learning, and showed a rapid convergence to the same execution time characteristics as the rote procedure condition.
One of the goals for the studies was to demonstrate that more complete and comprehensive mental model training would be superior to the sketcher training used in the original studies. Thus, the best condition was predicted to be one in which subjects studied the same generic inference rules and strategy as in the original simulation model. But this turned out to be the worst condition! Overall, the results were that the more complete and elaborate the mental model training, the longer it took people to learn it and to learn how to apply it.

A second goal was to get more stable, reliable time data by having the subjects repeat the problems, rather than solve them just one time, as before. Most frustrating was that initially, the training conditions could be distinguished, but performance was poor, and as the subjects gained experience, performance improved substantially, but the training conditions became indistinguishable. Thus the only effects of the mental model training were in the first few trials, and on these trials, the performance was generally no better in these experiments than they had been with the original ones. The procedure training condition was superior on these initial trials, and remained at least equal to all other conditions thereafter.

In summary, these studies suggest that subjects in the mental model conditions quickly construct a procedure for operating the device; they do this by making inferences from the mental model training materials, which can be difficult to do depending on the amount, complexity, and abstractness of the material. On the other hand, subjects given the procedure directly are far better off. Hence, acquiring procedures from text is superior to inferring them from a mental model.

Scaled-up mental model study. Another study was done in which the effects originally observed with the control panel device were sought in a scaled-up version. A computer-based version of the control panel device system was defined and implemented. Instead of operating switches and observing indicator lights, the subjects typed in commands and viewed status information on a video terminal. The three training conditions compared were (1) rote procedure training; (2) additional training in the syntax and semantics of the command language, which is most like traditional computer training; and (3) additional training on the underlying dynamic characteristics of the system, which directly supported inferences of correct operating procedures for novel situations.

The results showed a moderately strong overall benefit of more complete training, but this effect, contrary to expectation, failed to be specific to the individual problem solving situations. Furthermore, in a post hoc analysis of the procedural content of the training and testing, it appeared that to a good first approximation, the time taken to execute the task was a function of how much procedural knowledge the subjects had acquired in their previous experience with the device. This very powerful effect apparently masked many of the benefits that training in system knowledge might have provided.

Conclusions on mental model training. The overall conclusion from these studies is that high-quality training on procedures is probably the most efficient training approach. We have yet to see a case where training in mental model content is of definite benefit above and beyond high-quality procedure training. Perhaps there are situations where good procedure training is impractical. But otherwise, these studies imply that the benefits of understanding how a system works will be limited to very transitory first-time or one-of-a-kind situations. Once a subject has successfully problem-solved through a situation, the mental model content is of little value thereafter. These results are being prepared as a technical report.

Certainly further work on the value of mental models in training is needed, but we should probably be prepared for the outcome of this research being very different from what we would have previously imagined. For example, it could be that most mental model training in military equipment systems is simply a waste of time, given that in the actual work situation the procedures for interacting with most equipment have been made explicit, and when these procedures are not adequate to cover the situation, there is little that can be done other than to wait for an expert to repair the equipment. An example of this appears
in some of the actual Navy training and job aid materials that I have examined, in which even emergency malfunction procedures are spelled out in considerable detail. In contrast, the training includes classroom and textbook principles of systems such as steam turbines that are dwelt on at some length, even though nowhere in the trainee's job (or several grades above it) could this knowledge ever possibly be applied. Perhaps the major effect of such training is only to convey terminology and a knowledge of system components (e.g. types of valves and bearings) and certain elementary basic principles, such as why bearings can not be allowed to get too hot. Thus future work on this topic might best be focussed on documenting the redundancy and insignificance of much mental model content for many military jobs, rather than further attempts to demonstrate its usefulness.

**Acquiring Procedures from Text**

This work was carried on mostly by Susan Bovair. Extensions of this work and the simulation modeling of the data will be the subject of her dissertation.

**Major papers.** Two major papers on this topic were published with Bovair first author on both; the first was accepted at *Human-Computer Interaction*, with the data originally collected under IBM sponsorship. ONR support for the preparation of this paper was acknowledged. This paper concerned the learning and execution of procedures for operating a text editor. The second paper was an important theoretical paper on the acquisition of procedures from text for the new *Handbook of Reading Research*. This is a key paper because it is the first time a theoretically sophisticated treatment of this topic has been prepared and presented in a place where a variety of reading researchers will see it. One function of the paper is to emphasize the lack of research in this area; it is drastically under-explored. The paper also presents our framework for a model of how procedures are acquired, based on the construction of production rule representations for procedures using the information presented in the text. The preparation of this paper was an important activity in this project.

**“Overload” study.** Susan Bovair conducted an experiment on the “overload effect” noted in our earlier work on procedure acquisition. In this earlier experiment (Kieras and Bovair, 1986) we observed that one of the procedures, when it was the first one learned, took much more time than was predicted by the production rule model. Our hypothesis was that this was due to a working memory overload during procedure learning; this one procedure when learned first involved a total number of production rules that was substantially larger than that in any other situation in the experiment. If a representation of the entire procedure had to be constructed in working memory, then subjects presumably would have to engage in considerable extra processing to make up for the fact that they could not maintain the entire procedure at once in memory. This hypothesis was tested with a very carefully designed experiment that manipulated the degree of overload with a complex set of procedures defined on the control panel device. The experiment was designed using the simulation model of procedure acquisition to predict the number of new rules and overload effect *a priori*. A set of procedures and a set of training orders was defined so as to produce a predicted overload effect under conditions where it would not be confounded with other factors, such as being the very first trial in the experiment.

The basic results were that the putative overload effect occurred, but only early in the experiment; no overload effect occurred in the second half of the procedures. This initially very puzzling result fell into place when, as a result of the thinking involved in preparing the review paper, Bovair realized that there are a whole set of low-level procedures that subjects would have to learn in the experiment. For example, if one of the steps in the procedure was *Set S2 to X* then the subjects had to learn a whole set of production rules for carrying out what had appeared to us to be a single step. For example, first they have to locate the control S2, then grasp it, then rotate it in one direction or the other, and stop when the pointer is at the position X. Thus, there is in fact a set of lower-level production rules that must be learned.
in order to operate the equipment. Since these same basic activities are used repeatedly in all of the procedures, they only have to be learned early in the experiment.

Bovair was able to describe these low-level procedures and make quantitative predictions of training time using this expanded set of production rules. The fit to the data is very good. She has reanalyzed our earlier data from Kieras and Bovair (1986) with similar results. These results are also being prepared as a technical report.

Conclusions on procedure acquisition. The work done in this project on procedure acquisition has been largely theoretical, but with some important empirical results that show that the basic approach of modeling procedure learning in terms of the acquisition of production rules is both more robust and simpler than originally expected. Bovair's dissertation work, in which she will construct a comprehensive simulation model of the procedure acquisition process, should be a key piece of work in capturing the theoretical insights that we have accumulated thus far.

Comprehensibility System

Several versions of the comprehensibility system were delivered to NPRDC. This system has evolved considerably during its tenure as part of this project. The parser is now very sophisticated and very fast, and the set of criticism rules incorporated in the model have expanded into a fairly large and comprehensive set. This software was also repeatedly revised and improved both for efficiency and speed, but also for ease of future revision and updating. Additionally, the software was designed so that portions of it could be easily used in cognitive modeling projects such as Bovair's dissertation work. A chapter on the system appeared in the key Britton and Glynn volume on computer writing environments, ensuring its visibility to a wider audience.

At this point the major remaining problem with the comprehensibility system is that its output is still too verbose. The last work to be done on this version is to complete final work in extensions to the grammar, and simplify the output. One more version of the system is slated to be delivered to NPRDC, along with a technical report that provides maintenance and extension information for NPRDC staff.

Reports and Publications

Technical Reports


**Reports In Preparation**


**Publications**


**Problems Encountered**

The productivity of this project was impaired by many administrative and managerial problems. They are summarized here by way of explanation, and in hopes that they might be instructive for future projects. This project was initiated while I was still at the University of Arizona; then the project was moved to the University of Michigan. Of course there was the normal disruption due to the various problems of moving. But an especially serious one was that a new laboratory computer, a VAX 730, was installed. Although supplied at no cost to the project, considerable project resources during the first year were consumed in getting software converted and implemented. Given the obsolescence of the machine at this time, this large effort is not likely to pay off in the long run. It appears at this time an almost ideal environment for training experiments is a large screen Macintosh II with software packages such as HyperCard or Course of Action, which reduce the software programming requirements to a very low level, even though far more features and power are available.
Another computer-related problem was that at the time of the move, we had just received a set of Xerox 1108 Al Workstations, and a substantial effort was made to implement all cognitive modeling software on these machines and pursue other related work such as the comprehensibility system in this environment as well. These were very difficult machines to use, and consumed a very large amount of time. However, our actual computational needs were much simpler; we simply needed a high performance LISP implementation, and many of the features of the workstations that made them useful for intelligent tutoring system work were simply a distraction for us. A very large amount of software work was discarded after these machines were replaced with general purpose workstations (Apollos) which have proved to be far more satisfactory.

The most serious problem was a massive over-commitment of the Principal Investigator. In addition to the complications of moving to a different university and dealing with new lab equipment, I had too many projects active during the period of this project. In addition to this project, I was P.I. on a major IBM-sponsored project, a medium-sized project sponsored by NASA, the original comprehensibility system project, an additional ONR project that started later, and a substantial applied project for NPRDC. This over-commitment resulted in a lack of focus and erratic management of the staff. The most serious manifestation was that I had to take over the programming for the comprehensibility system. A more substantive effect of this over-commitment is that there were many promising research leads that opened up during this project that simply could not be followed up.

Accomplishments

Two items to mention under this heading involved aspects of this work making a transition towards application. The first was that the research on mental models was directly applied to work done for NASA on diagrammatic displays for engineered systems. The basic thrust was that diagrams, and especially computer-generated animated and color-coded diagrams, can be used to convey mental model information to the user. If this is done in a form that directly facilitates the inferences that the user needs to perform, problem solving ability with a system should be improved. These effects, along with some important limitations, were demonstrated both for the simple control panel device, and also for a complex piece of actual spacecraft equipment. The following is the reference for the report on this topic:


At the request of Dr. Gerald Laabs of NPRDC, we conducted a set of analyses of a test battery for job performance. These tests consisted of actual procedure executions for a variety of tasks involving ship machinery. Our analysis dealt with both the procedural and mental model content of the tasks and their relationship to the background knowledge and training of the personnel. The result was a suggestion for simple methods for ensuring non-redundant task selection. The reference for the report on this project is as follows:

References


