Technical Report: NAVTRASYSCEN TR86-002

CAI EVALUATION CHECKLIST: HUMAN FACTORS
GUIDELINES FOR THE DESIGN OF
COMPUTER-AIDED INSTRUCTION
Technical Report: NAVTRASYSNEN TR86-002

CAI EVALUATION CHECKLIST: HUMAN FACTORS
GUIDELINES FOR THE DESIGN OF
COMPUTER-AIDED INSTRUCTION

Cheryl J. Hamel
Stacie L. Clark
Human Factors Division
Naval Training Systems Center

August 1986

Final Report

DoD Distribution Statement
Approved for Public Release: distribution is unlimited
Human Factors design guidelines for the development of computer-aided instruction (CAI) are suggested. A checklist for CAI evaluation is provided. The discussion focuses on user-computer interaction in CAI. The guidelines are organized around five human factors design principles for computer systems: brevity, consistency; flexibility; responsiveness; and compatibility. Research plans are proposed to validate these guidelines and checklist and improve the scoring and diagnostics feedback features.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Purpose</td>
<td>2</td>
</tr>
<tr>
<td>II DEVELOPMENT OF THE CHECKLIST</td>
<td>3</td>
</tr>
<tr>
<td>Basic Principles of Good Computer System Design</td>
<td>3</td>
</tr>
<tr>
<td>Brevity</td>
<td>3</td>
</tr>
<tr>
<td>Consistency</td>
<td>3</td>
</tr>
<tr>
<td>Flexibility</td>
<td>4</td>
</tr>
<tr>
<td>Compatibility</td>
<td>4</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>5</td>
</tr>
<tr>
<td>Organization of the Guidelines</td>
<td>5</td>
</tr>
<tr>
<td>Scoring Method</td>
<td>5</td>
</tr>
<tr>
<td>Rater Reliability</td>
<td>6</td>
</tr>
<tr>
<td>Automated Checklist</td>
<td>7</td>
</tr>
<tr>
<td>III CONCLUSIONS AND RESEARCH RECOMMENDATIONS</td>
<td>8</td>
</tr>
<tr>
<td>Checklist Improvements</td>
<td>8</td>
</tr>
<tr>
<td>Diagnostic Feedback</td>
<td>9</td>
</tr>
<tr>
<td>Empirical Validation</td>
<td>9</td>
</tr>
<tr>
<td>Military Specifications</td>
<td>9</td>
</tr>
<tr>
<td>Conclusions</td>
<td>10</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>11</td>
</tr>
<tr>
<td>APPENDIX A--DESIGN GUIDELINES AND ASSOCIATED REFERENCES</td>
<td>15</td>
</tr>
<tr>
<td>APPENDIX B--CAI EVALUATION CHECKLIST</td>
<td>21</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BREVITY Items and Associated References</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>CONSISTENCY Items and Associated References</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>FLEXIBILITY Items and Associated References</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>COMPATIBILITY Items and Associated References</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>RESPONSIVENESS Items and Associated References</td>
<td>20</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

The authors wish to thank Drs. Richard Braby and Mark Pfeiffer from the Naval Training Systems Center, who contributed their technical assistance in support of this effort. Dr. Braby, an education specialist, reviewed an initial version of the checklist for clarity and relevance to computer-based instruction. Dr. Pfeiffer, a human factors psychologist, reviewed the checklist for its relevance to interactive videodisk training and collected data for the study of rater reliability.
INTRODUCTION

The Navy is cautiously expanding the use of computer-aided instruction (CAI) in classrooms and embedded training in operational environments. If the effectiveness of such instruction is to be maximized, attention must be given to the user-computer interface (UCI). The ease with which students communicate with the computer, in part, will determine how well they learn. In the case of embedded training, the UCI may determine whether the training is used at all.

Learning complex job tasks consumes students' attention and memory resources. Additional demands simply to understand the complex system serving as the medium of instruction should be minimized. Design of CAI according to human factors principles will lead to a good match between cognitive processes of students and the information-processing capabilities of computers. The result will be instruction which uses all the advantages of the computer as a training medium to produce high information transfer, reduced mental workload, reduced human error and efficient learning.

Research articles and guidelines on the design of CAI and on the UCI have proliferated in recent years. These studies and guidelines have not been organized and evaluated for their relevance to training effectiveness.

The Naval Training Systems Center is conducting research to produce human factors guidelines for the design and evaluation of computer-based training systems. The first product of this research effort, reported here, is a set of draft guidelines, with accompanying references, for the design of CAI. These guidelines were derived from literature reviews, existing guidelines and standards, and empirical and theoretical research pertaining to CAI and the UCI. Special attention has been given to information-processing theory which focuses on the cognitive demands placed upon an operator of a computer system.

The guidelines presented in this report are organized according to a set of human-computer dialogue principles proposed by Williges and Williges (1984) and are formatted as the CAI Evaluation Checklist. The checklist format was chosen to examine its potential for evaluating the interaction of a student with a computer. Additional research will be conducted to validate the instrument's ability to predict the training effectiveness of CAI. Future research will examine the checklist's ability to measure the quality of CAI courseware and hardware, in terms of how well a system displays information, how easy it is for students to
operate a system, and how well the CAI capitalizes on the advantages of the computer as a training medium.

PURPOSE

This report provides guidelines in checklist format for the design of the user-computer interface (UCI) for computer-assisted instruction (CAI). As such, it will be of interest to human factors professionals, instructional technologists and engineers who design or develop computer-based training systems or authoring systems for CAI.

The checklist format is useful for evaluating CAI as part of the final acceptance process for the government acquisition of CAI systems and also to those doing research on CAI or the UCI who need quantitative methods of evaluation.
The guidelines were produced from a review of the behavioral research literature, existing guidelines, and verbal reports of experienced CAI developers. Information on the design of CAI was combined with studies of the UCI in an effort to merge the two areas. Appendix A includes the guidelines and specific references from which they were derived.

BASIC PRINCIPLES OF GOOD COMPUTER SYSTEM DESIGN

The Williges and Williges (1984) report on human factors consideration in the design of human-computer dialogue provided an organizational structure for the material. All guidelines were organized using five of the Williges' principles: brevity, consistency, flexibility, compatibility, and responsiveness. From the Williges' (1984) review of existing standards and guidelines on the human-computer interface, they concluded that these principles are fundamental to "good computer system design." The definition of each principle was modified to fit the requirements of CAI as a particular form of human-computer interaction. The modified principles are described below.

BREVITY

Information a student must maintain in short-term memory or attend to should be minimized. Methods of chunking and focusing information aid in reducing mental workload. Early research on information processing in humans supports this principle (Miller, 1960; Norman, 1976). Humans have a limited capacity to process quantities of information at one time and therefore must selectively attend or use other control processes to receive and remember information.

CONSISTENCY

Task demands must be consistent within a training system to develop user expectations. If a student develops a correct mental model of the system, there will be a dramatic reduction in cognitive processing expended on understanding the computer and how it works.
Recent studies invoke the concept of mental models to explain skilled interaction with a system. These models explain learning as the development of successive mental representations of the system and the acquisition of a set of rules for interacting with that system (Card, Moran, and Newell (1980). With repetition of particular combinations of actions, cognitive processing becomes proceduralized, or automatic (Anderson, 1981).

Theories on the development of automatic skills stress the importance of training the consistent components of the training tasks (Schneider and Shiffrin, 1977).

FLEXIBILITY

A system, including the instruction, must accommodate individual differences among students. A computer system must have the capabilities for review and branching, and the instruction should include skill diagnosis and remediation. Flexibility also includes capabilities for student control over the instruction, such as pacing.

Individual differences is a central concern in human learning and performance. The cognitive processes (e.g., memory operations such as storing, searching and retrieving) that intervene during learning and performance are individual difference variables to be considered in designing adaptive instruction (Glaser & Resnick, 1972; Hunt & Lansman, 1975).

COMPATIBILITY

To minimize the information processing between stimulus and response, input and output formats must be compatible with each other and with established behavior patterns of the students using the system. There must be a minimum of translation, decoding, and other forms of cognitive processing necessary to understand computer output of information and to know what response (input) is required.

First, how information is presented must be compatible with the required format for student responses. When tasks are designed such that the response mode is cognitively compatible with the presentation mode, performance is faster and more accurate than when incompatibility exists. Stimulus-Central processing-Response (S-C-R) Compatibility theory and related empirical research support these contentions (Wickens, Sandry, & Vidulich, 1983). The theory maintains that when a stimulus is presented, central processing of that stimulus will either be verbal or spatial in nature. Verbal responses are cognitively compatible with verbal processing tasks and motor responses with spatial processing tasks.
Second, input and output must conform to population stereotypes, e.g., red=stop, green=go. If cues used in the training material must be encoded in memory such that they conflict with previous encoding of those cues, the result will be increased information processing (interference) during retrieval from memory (Tulving, 1976).

RESPONSIVENESS

The system must provide informative feedback about student performance and system condition. Optimal timing of system responses to student input will help the students to know where they are, what they have done, whether or not it was successful, and how to take corrective action as necessary.

Research indicates that feedback which provides information, not simply immediate feedback, is the key to performance change (Cohen, 1985). Knowledge of results in the form of immediate feedback like "CORRECT" or "ERROR," are often not enough to correct the problem. Informational feedback helps the student locate the error and construct an alternative response, e.g., "The first part of your answer is not valid, check your arithmetic."

ORGANIZATION OF THE GUIDELINES

The authors used the five principles described above to sort 50 guidelines into five categories, and then obtained expert opinion on the assignment of guidelines to the categories. The categorized checklist items were distributed to one specialist each in computer software, education, and human factors for their review and comments. They provided constructive comments on the clarity and importance of items and their appropriateness to a given category. The author used these comments to make appropriate modifications to the checklist.

The unique categorization opened up possibilities of measuring the quality of the UCI through the measurement of five characteristics—brevity, consistency, flexibility, compatibility, and responsiveness. A simplified scoring method was derived to begin a study of these possibilities.

SCORING METHOD

The number of guidelines in each category was kept to a maximum of ten items. A yes/no format was used to make the checklist easy to use. The user is asked to mark "yes" if the system contains the feature for the most part, and "no" if the system rarely or never incorporates this
feature. The "n/a" response was added for those cases where an item does not apply. The items were given equal weighting since a weighting method is not yet developed. A score in each category can be obtained by the following formula:

\[
\text{Score} = \frac{\text{Number of "yes" responses}}{\text{Total number of items} - \text{"n/a" responses}} \times 100\%
\]

This scoring method results in scores for each category ranging from 0% to 100%. Criterion scores for acceptable design of computer-aided instruction, currently unknown, will be developed and refined through empirical research.

RATER RELIABILITY

In an initial attempt to obtain rater reliability, education and psychology professionals were asked to use the CAI Evaluation Checklist to evaluate two articles of interactive video courseware: Trace-A-Ground and AC Switchboard developed for the Electrician's Mate Navy "A" school. The content was introductory material for electronics training.

Four evaluators viewed the Trace-A-Ground courseware and five evaluators viewed the AC Switchboard material. Rater reliability was assessed using the analysis of variance approach to correlation (Winer, 1962). The scores on each of the five principles were set up for a 2 x 2 ANOVA with the principles as one factor and judges as the other factor. Two analyses were performed, one for each set of materials. The obtained mean squares were substituted into the Spearman-Brown prophecy formula to obtain estimated reliability of the averaged scores on the five principles for two, three, four, and five raters. The first ANOVA showed that the reliability of the average of the four raters who evaluated Trace-A-Ground was r=0.87. This means that an average score on brevity, consistency, flexibility, compatibility, and responsiveness from four raters will correlate r=0.87 with the averages obtained from a second set of four raters. The second ANOVA showed the reliability of the average of the five raters who evaluated AC Switchboard to be r=0.91. The Spearman-Brown prophecy formula indicated that acceptable reliability (.80) can be achieved with only two raters. Further information can be obtained from Pfeiffer, Miller, Platt, Green, Monroe, and Traxler (1986).
For the purposes of simplifying data collection, the CAI Evaluation Checklist has been programmed in BASIC on a TRS-80 Model 200 lap-top portable computer. The computer weighs about 4 lbs. and can be battery-operated. The machine uses a 40 x 16 character LCD screen and has a full-size "QWERTY" keyboard. Data storage is in the 72k byte RAM and additional portable disk drives are available. Programs are available to interface the Model 200 with other microcomputers so that files can be written directly to disk.

The automated checklist is useful for large-scale data collection and data analysis. Data are stored for each rater and the averages of all raters are computed. Plans are to program computations for obtaining reliability scores, and to program diagnostic feedback. As an example, the diagnostic feedback will contain information on how best to revise the materials considering what changes should have the highest priority. Prioritizing will be possible when further research produces a method for weighting individual checklist items.
SECTION III

CONCLUSIONS AND RESEARCH RECOMMENDATIONS

The goal of subsequent research in this project is to provide human factors professionals, engineers, and education specialists with a valid and reliable checklist for computer-assisted instruction evaluation. An additional effort could produce military specifications from the checklist items.

The following areas will receive attention in continuing research to determine the validity of the checklist:

1) Addition of guidelines to the item pool, and determination of the appropriateness of guidelines to a category,
2) Improvement of the scoring method by the weighting of items in each category,
3) Interpretation of scores and diagnostic feedback through automation.
4) Empirical tests of the checklist's ability to predict training effectiveness.

CHECKLIST IMPROVEMENTS

Further review of the literature and on-going research will add to the data base on the human-computer interface and computer-assisted instruction. The reviews may produce additional items for the checklist.

To determine the categorization of items, human factors experts, software engineers, and education specialists will be asked to categorize the items using the five dimensions described earlier (Williges and Williges, 1984). Items will be shifted to another category if there is high agreement among the experts that a change is warranted. Items not reliably assigned to a category will be revised or dropped from the checklist.

A new method of scoring based on weighted items will be developed. After items have been categorized, the Likert scaling technique will be used to weight each item within the categories. Experts will be asked to rate each item on its importance for effective learning. The weighted scoring method will increase the capability of the checklist to predict
training effectiveness, and will contribute to more accurate diagnostic feedback.

The Likert-scaling format will replace the yes-no format of the checklist if research indicates improvement in the instrument's validity.

DIAGNOSTIC FEEDBACK

Weighted checklist items will prioritize for the user those items most significantly related to training effectiveness. If scores are entered on the automated version of the checklist, it will be possible to obtain automated diagnostic feedback. This feedback will interpret each of the five scores in terms of what revisions are most crucial for increasing the quality of the CAI.

EMPIRICAL VALIDATION

Empirical research will be conducted to evaluate the potential of the CAI Evaluation Checklist to predict training effectiveness. Over 50 government agencies and private firms involved in the development or evaluation of computer-assisted instruction have been given a copy of the CAI Evaluation Checklist. Data will be obtained on the applicability of the checklist to various types of CAI (e.g., tutorial, drill and practice, simulation, games). If measures of student performance are available, or other evaluative measures, they will be reported along with checklist measures.

The next major research effort will focus on experimental designs to test the ability of the checklist to predict training effectiveness. One research plan is to compare two versions of CAI with the same instructional content, one designed with the checklist guidelines and one without. The two CAI versions will be rated independently with the checklist after all the design and development have been completed. The two CAI versions will then be compared on training effectiveness.

Using a between subjects design, students will learn one of the two versions and statistical techniques will be used to correlate student performance with the CAI Evaluation Checklist measures.

MILITARY SPECIFICATIONS

Another research effort is to rewrite the design guidelines in the form of military specifications. A team of education specialists involved in front-end analysis and human factors experts will use the checklist items as the basis for a Data Item Description (DID). The DID could accompany contractual agreements with the Naval Training Systems Center.
when the item to be delivered is computer-based instruction or interactive videodisk instruction.

CONCLUSIONS

The CAI Evaluation Checklist has been derived from research and guidelines on the design of the human-computer interface for computer-aided instruction. The guidelines have been formatted as a checklist to be used as an evaluation tool. Proposed research should improve the validity of the checklist and improve the scoring procedure.

A checklist that is easy to administer, has high face validity, and is correlated with real world performance should gain wide acceptance as an extremely cost-effective evaluation tool.

It is proposed that the checklist items be rewritten as military specifications for computer-based instruction delivered to the Naval Training Systems Center.
REFERENCES


Table 1. BREVITY items and associated references.

<table>
<thead>
<tr>
<th>BREVITY ITEM</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large portions of text are broken into meaningful &quot;chunks&quot;.</td>
<td>Caldwell, 1980; Kearsley and Hillelsohn, 1982</td>
</tr>
<tr>
<td>No more than seven lines of text per screen.</td>
<td>Gilmore, 1985; Karlsrud, 1985</td>
</tr>
<tr>
<td>Graphic displays activate from 15% to 25% of the screen area.</td>
<td>Tullis, 1983</td>
</tr>
<tr>
<td>Main menus and sub-menus have 3 to 9 choices.</td>
<td>Lee and McGregor, 1985; Tijerina, Chevalaz, and Meyers, 1985</td>
</tr>
<tr>
<td>Use of color, boxing, and highlighting, rather than blinking to focus attention on important segments of text.</td>
<td>Caldwell, 1980; Williges and Cheleaza, 1984; Belezza and Cheney, 1973</td>
</tr>
<tr>
<td>No more than 3 or 4 text screens without interactivity.</td>
<td>Mahoney and Lyday, 1984</td>
</tr>
<tr>
<td>No more than 10% of the screen display is highlighted at one time.</td>
<td>Gilmore, 1985</td>
</tr>
<tr>
<td>The time required for a typical session (or lesson) is within the attention span of the target audience.</td>
<td>Kearsley and Hillelsohn, 1982</td>
</tr>
<tr>
<td>Sentences have simple syntax: active voice, not compounded.</td>
<td>Boyd and Eldridge, 1984; Glynn and Britton, 1984</td>
</tr>
</tbody>
</table>
Table 2. CONSISTENCY items and associated references.

<table>
<thead>
<tr>
<th>CONSISTENCY ITEM</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functionally alike screens are formatted in the same way.</td>
<td>Sawyer, 1985; Williges and Williges, 1984</td>
</tr>
<tr>
<td>When functional areas are erased, they are consistently rewritten in the same order.</td>
<td>Heines, 1984</td>
</tr>
<tr>
<td>Consistent use of labels and graphics keeps the same types of frames identified as such.</td>
<td>Gilmore, 1985; Military Standard MIL-STD-1472C, 1981</td>
</tr>
<tr>
<td>Critical information is always presented at the beginning of a message or centered on the screen.</td>
<td>Gilmore, 1985; Swezey and Davis, 1983</td>
</tr>
<tr>
<td>Students receive constant delay of feedback (no more than 2 seconds) rather than variable delays.</td>
<td>Kearsley and Hillelsohn, 1982; McCaan, 1983</td>
</tr>
<tr>
<td>Consistency in the way questions are asked and the required format for responses.</td>
<td>Sawyer, 1985; Williges and Williges, 1984</td>
</tr>
<tr>
<td>The structure of the presentation is evident to the user through the use of menus and concepts maps.</td>
<td>McCaan, 1983; Snyder, Happ, Malcus, Paap, and Lewis, 1985</td>
</tr>
<tr>
<td>A symbol has the same meaning at all times.</td>
<td>Gilmore, 1985</td>
</tr>
</tbody>
</table>
Table 3. FLEXIBILITY items and associated references.

<table>
<thead>
<tr>
<th>FLEXIBILITY ITEM</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>A page-back capability allows the student to review previous material.</td>
<td>Caldwell, 1980</td>
</tr>
<tr>
<td>Students can easily exit lessons, return to menus, and exit the program.</td>
<td>Caldwell, 1980; Mahoney and Lyday, 1984</td>
</tr>
<tr>
<td>Student has control over the rate of presentation of frames.</td>
<td>Caldwell, 1980</td>
</tr>
<tr>
<td>Flexibility in recognizing correct responses (e.g., misspellings, partial answers).</td>
<td>Caldwell, 1980; Kearsley and Hillelsohn, 1982; McPherson-Turner, 1979</td>
</tr>
<tr>
<td>The student can request more lengthy messages, through helps, if further clarification is needed.</td>
<td>Gilmore, 1985</td>
</tr>
<tr>
<td>The program contains activities for diagnosis of skills already mastered.</td>
<td>Caldwell, 1980</td>
</tr>
<tr>
<td>There are remedial exercises for skill deficiencies.</td>
<td>Caldwell, 1980</td>
</tr>
<tr>
<td>Modularized program (with menus) allows the student to begin at a point appropriate to past achievement.</td>
<td>Caldwell, 1980; Sawyer, 1985</td>
</tr>
<tr>
<td>The student can choose the difficulty level of problems or exercises (achieved by variation in the use of prompts).</td>
<td>Boyd and Eldridge, 1984; Caldwell, 1980; Sawyer, 1985</td>
</tr>
<tr>
<td>The student can correct an input error (e.g., with BACKSPACE) or recover from input errors without disrupting the lesson sequence.</td>
<td>Williges and Williges, 1984</td>
</tr>
</tbody>
</table>
Table 4. COMPATIBILITY items and associated references.

<table>
<thead>
<tr>
<th>COMPATIBILITY ITEM</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>The response mode is appropriate to the target audience.</td>
<td>Williges and Williges, 1984</td>
</tr>
<tr>
<td>Students are required to use codes for responding only when necessary, as in multiple choice answering (e.g., 1=yes, 2=no is unnecessary coding).</td>
<td>Caldwell, 1980</td>
</tr>
<tr>
<td>Visual information and visual tasks such as locating or repositioning are presented graphically.</td>
<td>Wickens, Sandry, and Vidulich, 1983</td>
</tr>
<tr>
<td>Where frames are labeled, title not number is used for identification.</td>
<td>Boyd and Eldridge, 1984</td>
</tr>
<tr>
<td>If commands are entered by keyboard, the student types the first letter rather than numeric code (e.g., h=help).</td>
<td>Williges and Williges, 1984; Shinar, Stern, Bubis, and Ingram, 1985</td>
</tr>
<tr>
<td>Input, output is consistent with user population stereotypes (e.g., correct (e.g., correct response feedback in green).</td>
<td>Barhard, Hammond, Morton, and Long, 1981</td>
</tr>
<tr>
<td>Where order of lessons is important, menu options are listed by number not by letter.</td>
<td>Tijerina, et al., 1985</td>
</tr>
<tr>
<td>To clarify drill or test instructions, a sample item is answered before the drill or quiz begins.</td>
<td>Caldwell, 1980</td>
</tr>
<tr>
<td>A response is demanded while the instructions on how to respond are still on the screen.</td>
<td>Caldwell, 1980</td>
</tr>
<tr>
<td>Routing menus are limited to a maximum of three levels.</td>
<td>Heines, 1984</td>
</tr>
</tbody>
</table>
Table 5. RESPONSIVENESS items and associated references.

<table>
<thead>
<tr>
<th>RESPONSIVENESS ITEM</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>When the student must stand by, periodic feedback indicates normal operation.</td>
<td>Military Standard MIL-STD-1472C, 1981; Williges and Williges, 1984</td>
</tr>
<tr>
<td>The computer tracks response patterns and lists areas where the student remediation.</td>
<td>Cohen, 1985; Mahoney and Lyday, 1984</td>
</tr>
<tr>
<td>Feedback and directions are clearly distinguishable from other text through use of color, boxing, reverse video, etc.</td>
<td>Williges and Williges, 1984</td>
</tr>
<tr>
<td>Students can obtain a score.</td>
<td>Cohen, 1985; McPherson-Turner, 1979</td>
</tr>
<tr>
<td>At higher mastery levels, students are given immediate knowledge of right and wrong responses, and more lengthy feedback is delayed until the end of the session.</td>
<td>Cohen, 1985</td>
</tr>
<tr>
<td>There is a pause after feedback, before the lesson continues, to allow time for consolidation of the newly acquired material.</td>
<td>Kearsley and Hillelsohn, 1982; Ramsey and Atwood, 1979</td>
</tr>
<tr>
<td>Access to helps, references, or resources is easily available.</td>
<td>Gilmore, 1985</td>
</tr>
<tr>
<td>At the beginning of training, feedback is response informative (e.g., the part of your answer is incorrect.)</td>
<td>Caldwell, 1980</td>
</tr>
<tr>
<td>The student gets more than one chance to give the answer (with prompts).</td>
<td>Caldwell, 1980; Military Standard MIL-STD-1472C, 1981</td>
</tr>
<tr>
<td>It takes no more than 5 seconds for text and graphics to fill the screen.</td>
<td>Gilmore, 1985; Kearsley and Hillelsohn, 1982</td>
</tr>
</tbody>
</table>
APPENDIX B

CAI EVALUATION CHECKLIST
HUMAN-COMPUTER INTERFACE DESIGN BASED UPON INFORMATION PROCESSING PRINCIPLES

by

Cheryl J. Hamel
Human Factors Division
Naval Training Systems Center

March 26, 1986

Presented at the Southeastern Psychological Association Meeting, Kissimmee, Florida as part of a symposium entitled, "Emerging Solutions to Problems in the Use of Computer-Assisted Instruction for Military Training."
The CAI Evaluation Checklist examines the user-computer interaction involved in computer aided instruction. The checklist is based on five human factors design guidelines for computer systems proposed by Williges and Williges (1984): brevity, consistency, flexibility, responsiveness and compatibility.

High scores on these factors purport to lead to high information transfer, low mental workload, and reduction of human error.
If you use this checklist to evaluate CAI courseware, please return the following information to me.

1. Name of your organization ____________________________________________

2. Name of CAI course or lesson __________________________________________

3. Type of CAI (You may check more than one.)

   Drill and practice ____  Game ____
   Tutorial ____  Intelligent CAI ____
   Simulation ____  Embedded Training ____

4. Rate the courseware based on your own expert opinion:

   poor  average  very good

5. Record the scores you obtained from the checklist:

   yes  no  n/a

   Brevity  ____  ____  ____
   Consistency  ____  ____  ____
   Flexibility  ____  ____  ____
   Compatibility  ____  ____  ____
   Responsiveness  ____  ____  ____

Return this page to:

Cheryl J. Hamel, Ph.D.
Human Factors Division, Code 711
Naval Training Systems Center
Orlando, FL 32813
INSTRUCTIONS

For each item on the checklist, check:

"yes" if the instructional system contains this feature, for the most part.

"no" if the instructional system rarely or never incorporates this feature.

"N/A" if the item does not apply.
CONSISTENCY

In a consistent system, the user will develop a conceptual model of how the system works and will gain confidence.

<table>
<thead>
<tr>
<th>yes</th>
<th>no</th>
<th>n/a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Functionally alike screens are formatted in the same way.
2. When functional areas are erased, they are consistently rewritten in the same order.
3. Consistent use of labels and graphics keeps the same types of frames identified as such.
4. Critical information is always presented at the beginning of a message or centered on the screen.
5. Students receive constant delay of feedback (no more than 2 seconds) rather than variable delays.
6. Consistency in the way questions are asked and the required format for responses.
7. The structure of the presentation is evident to the user through the use of menus and concepts maps.
8. A symbol has the same meaning at all times.
COMPATIBILITY

The input-output format focuses attention on the task to be learned, not on the task of learning the system.

yes no n/a

1. The response mode is appropriate to the target audience.

2. Students are required to use codes for responding only when necessary, as in multiple choice answering (e.g., 1=yes, 2=no is unnecessary coding).

3. Visual information and visual tasks such as locating or repositioning are presented graphically.

4. Where frames are labeled, title not number is used for identification.

5. If commands are entered by keyboard, the student types the first letter rather than numeric code (e.g., h=help).

6. Input, output is consistent with user population stereotypes (e.g., correct response feedback in green).

7. Where order of lessons is important, menu options are listed by number, not by letter.

8. To clarify drill or test instructions, a sample item is answered before the drill or quiz begins.

9. A response is demanded while the instructions on how to respond are still on the screen.

10. Routing menus are limited to a maximum of three levels.
BREVITY

Displays are simple and well-organized for high information transfer.

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>no</th>
<th>n/a</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Large portions of text are broken into meaningful &quot;chunks&quot;.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>No more than seven lines of text per screen.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Graphic displays activate from 15% to 25% of the screen area.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Main menus and sub-menus have 3 to 9 choices.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Use of color, boxing, and highlighting, rather than blinking, to focus attention on important segments of text.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>No more than 3 or 4 text screens without interactivity.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>No more than 10% of the screen display is highlighted at one time.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>The time required for a typical session (or lesson) is within the attention span of the target audience.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Sentences have simple syntax: active voice, not compounded.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FLEXIBILITY

The program adapts to individual differences among users.

yes no n/a

-- -- -- 1. A page-back capability allows the student to review previous material.

-- -- -- 2. Students can easily exit lessons, return to menus, and exit the program.

-- -- -- 3. Student has control over the rate of presentation of frames.

-- -- -- 4. Flexibility in recognizing correct responses (e.g., misspellings, partial answers).

-- -- -- 5. The student can request more lengthy messages, through helps, if further clarification is needed.

-- -- -- 6. The program contains activities for diagnosis of skills already mastered.

-- -- -- 7. There are remedial exercises for skill deficiencies.

-- -- -- 8. Modularized program (with menus) allows the student to begin at a point appropriate to past achievement.

-- -- -- 9. The student can choose the difficulty level of problems or exercises (achieved by variation in the use of prompts).

-- -- -- 10. The student can correct an input error (e.g., with BACKSPACE) or recover from input errors without disrupting the lesson sequence.
RESPONSIVENESS

Feedback is informative and timing is optimal.

yes no n/a

1. When the student must stand by, periodic feedback indicates normal operation.

2. The computer tracks response patterns and lists areas where the student needs remediation.

3. Feedback and directions are clearly distinguishable from other text through use of color, boxing, reverse video, etc.

4. Students can obtain a score.

5. At higher mastery levels, students are given immediate knowledge of right and wrong responses, and more lengthy feedback is delayed until the end of the session.

6. There is a pause after feedback, before the lesson continues, to allow time for consolidation of the newly acquired material.

7. Access to helps, references, or resources is easily available.

8. At the beginning of training, feedback is response informative (e.g., the --- part of your answer is incorrect.)

9. The student gets more than one chance to give the correct answer (with prompts).

10. It takes no more than 5 seconds for text and graphics to fill the screen.
SCORING

A score on each factor (brevity, consistency, flexibility, responsiveness and compatibility) can be obtained. To obtain a score on a factor:

1. Add the total number of "Yes" responses.
2. Subtract the number of "N/A" responses from the total.
3. Divide to obtain a percentage of "Yes" responses, and multiply by 100%.

Example: Brevity Score = \[
\frac{\text{Number of "Yes" responses}}{\text{Total}} \times 100\%
\]

\[
9 - \text{"N/A" responses}
\]
<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>Phone Number</th>
<th>Email</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commanding Officer</td>
<td>Naval Training Systems Center</td>
<td>Orlando, FL 32813-7100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edward A. Martin</td>
<td>Technical Advisor</td>
<td>ASD/ENETS</td>
<td>Wright-Patterson Air Force Base</td>
<td></td>
</tr>
<tr>
<td>LTC Mike McGaugh</td>
<td>PM Training Devices</td>
<td>ATTN: AMPCM-AVD</td>
<td>Orlando, FL 32813-7100</td>
<td></td>
</tr>
<tr>
<td>Naval Personnel Research and Development Center</td>
<td>ATTN: Russel M. Vorce, Code 31</td>
<td>San Diego, CA 92152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFHRL/FTR</td>
<td>ATTN: Robert S. Kellogg, Ph.D.</td>
<td>Bldg 588 (UDRI)</td>
<td>Williams Air Force Base</td>
<td>Chandler, AZ 85224-5000</td>
</tr>
<tr>
<td>Naval Aerospace Medical Institute</td>
<td>ATTN: Code 00L, COL F. S. Pettyjohn</td>
<td>Naval Air Station</td>
<td>Pensacola, FL 32505</td>
<td></td>
</tr>
<tr>
<td>Naval Training Systems Center</td>
<td>ATTN: Code 002, COL R. C. Baker</td>
<td>Orlando, FL 32813-7100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Patrick McCann</td>
<td>Human Factors, Code 71</td>
<td>Naval Personnel &amp; Research Laboratory</td>
<td>San Diego, CA 92152-6800</td>
<td></td>
</tr>
<tr>
<td>Naval Training Systems Center</td>
<td>Aviation Plans &amp; Analysis Branch</td>
<td>Code 111</td>
<td>ATTN: Brenda Hubbard</td>
<td>Orlando, FL 32813-7100</td>
</tr>
<tr>
<td>Air Force Human Resources Laboratory</td>
<td>ATTN: Mike Venturino</td>
<td>Wright-Patterson Air Force Base</td>
<td>Dayton, OH 45433</td>
<td></td>
</tr>
<tr>
<td>Defense Technical Information Center</td>
<td>Cameron Station</td>
<td>Alexandria, VA 22310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. G. Williams</td>
<td>Naval Training Systems Center</td>
<td>Code L02</td>
<td>Pensacola, FL 32508</td>
<td></td>
</tr>
<tr>
<td>Mike Frazier</td>
<td>AFTEC/TELH</td>
<td>Air Force Test &amp; Evaluation Center</td>
<td>Kirtland Air Force Base, NM 87717</td>
<td></td>
</tr>
<tr>
<td>Air Force Human Resources Laboratory</td>
<td>ATTN: Thomas H. Killion, Ph.D.</td>
<td>Williams Air Force Base</td>
<td>Chandler, AZ 85224-5000</td>
<td></td>
</tr>
<tr>
<td>Air Force Human Resources Laboratory</td>
<td>OT Division</td>
<td>Williams Air Force Base</td>
<td>Chandler, AZ 85224-5000</td>
<td></td>
</tr>
<tr>
<td>William J. Thomas</td>
<td>Fleet Analysis Center</td>
<td>Corona, CA 91720</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michael A. Vidulich</td>
<td>NASA - Ames Research Center</td>
<td>Moffett Field, CA 94035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gilbert L. Neal, Ph.D.</td>
<td>USA TRADOC Systems Analysis Activity</td>
<td>ATTN: ATOR-TH</td>
<td>White Sands Missile Range, NM 88002-5002</td>
<td></td>
</tr>
<tr>
<td>Thomas E. Ulrich, MAJ, USAF</td>
<td>Chief, Training Research Branch</td>
<td>Department of the Air Force</td>
<td>Randolph Air Force Base, TX 78150-5001</td>
<td></td>
</tr>
<tr>
<td>Dr. Gavin Lintern</td>
<td>Aviation Research Laboratory</td>
<td>University of Illinois</td>
<td>Willard, IL 61874</td>
<td></td>
</tr>
</tbody>
</table>
Mr. Chuck Gainer  
Chief, ARI, Field Unit- Ft. Rucker  
ATTN: PERI-SR  
Fort Rucker, AL  36362

CAPT Kevin Smith, USNR  
COMFITAEWINGPAC  
Naval Air Station  
Miramar, CA  92145

Dr. Alva C. Bittner, Jr  
U.S. Naval Biodynamics Laboratory  
Box 20189 Michoud Street  
New Orleans, LA  70189

LCDR Thomas Crosby, MSC,USN  
Naval Air Systems Command  
ATTN: 930J  
Washington, DC  20361

LT Dan Dolgin, MSC, USNR  
Operartional Psychology Dept  
Code 11  
Naval Air Station  
Pensacola, FL  32508

CAPT Thomas Gallagher, MSC, USN  
Naval Air Development Center  
Code 60A  
Warminster, PA  18974

Dr. Roger Whiteway  
Naval Strike Warfare Center  
Naval Air Station  
Fallon, Nevada  89406

LCDR Steve Harris, MSC, USN  
Naval Air Test Center  
Code SY70H  
Patuxent River, MD  20670

CDR Chuck Hutchins, MSC, USN  
Naval Post Graduate School  
Code 55MP  
Monterey, CA  93940

CAPT James Goodson, MSC, USN  
Operational Psychology Dept  
Naval Aerospace Medical Institute  
Code 11  
Naval Air Station  
Pensacola, FL  32508

Dr. John Chippendale  
PERI-SR  
Bldg 501  
For Rucker, AL  36362

Ray Cavallero  
Pacific Missile Test Center  
Code 1042  
Pt Mugu, CA  93042

CAPT Paul Chatelier, MSC, USN  
OUSDR&E (R&AT)  
The Pentagon (Room 3D129)  
Washington, DC  20361

LCDR Larry Frank, MSC, USN  
715 Broce Drive  
Blacksburg, VA  24060

CDR Wade Helm, MSC, USN  
Naval Aerospace Medical Research Laboratory (Code 05)  
Naval Air Station  
Pensacola, FL  32508

CAPT Joseph Funaro, MSC, USN  
Naval Training Systems Command  
Code 71  
Orlando, FL  32813-7100

LT James Hooper, MSC, USNR  
Naval Air Systems Command  
ATTN: APC205-ON  
Washington, DC  20361

LT Lee Goodman, MSC, USN  
Naval Air Development Center  
Code 6002  
Warminster, PA  18975

CDR Dennis McBride, MSC, USN  
Pacific Missile Test Center  
Code 4025  
Point Mugu, CA  93042

CDR Thomas Jones, MSC, USN  
Office of Naval Research  
Code 125  
800 N. Quincy Street  
Arlington, VA  22217
END

DTIC

10-86