Behavior Research Methods, Instruments, & Computers
2004, 36 (2), 241-249

Human use regulatory affairs advisor (HURAA): Learning about research ethics with intelligent learning modules

XIANGEN HU and ARTHUR C. GRAESSER
University of Memphis, Memphis, Tennessee

The Human Use Regulatory Affairs Advisor (HURAA) is a Web-based facility that provides help and training on the ethical use of human subjects in research, based on documents and regulations in United States federal agencies. HURAA has a number of standard features of conventional Web facilities and computer-based training, such as hypertext, multimedia, help modules, glossaries, archives, links to other sites, and page-turning didactic instruction. HURAA also has these intelligent features: (1) an animated conversational agent that serves as a navigational guide for the Web facility, (2) lessons with case-based and explanation-based reasoning, (3) document retrieval through natural language queries, and (4) a context-sensitive Frequently Asked Questions segment, called *Point & Query*. This article describes the functional learning components of HURAA, specifies its computational architecture, and summarizes empirical tests of the facility on learners.

Many Web-based learning environments of the future will be general-purpose knowledge acquisition facilities rather than sites with narrowly targeted learning objectives. An adequate learning environment will hook the learner into using the site with motivating features, will guide him or her in covering relevant material through appropriate pedagogical methods, will answer a broad range of questions that students might ask, and will serve as an information repository. Such learning environments may be viewed as a help desk for active inquiry rather than, or in addition to, a structured regimen of information delivery and testing. The purpose of this article is to describe the general-purpose learning environment that we developed in a project funded by the Department of Defense (DoD). The system is called HURAA, an acronym for *Human Use Regulatory Affairs Advisor*.

HURAA is a Web-based facility that provides help, training, and information retrieval on the ethical use of human subjects in research. The content of HURAA is derived from federal agency documents and regulations, particularly the National Institutes of Health (NIH 45 CFR 46), DoD (32 CFR 219; Directive 3216.2), and particular branches of the US military. The users of HURA are high-ranking military officials who must approve research protocols involving human subjects in studies sponsored by DoD. These users focus on fundamental ethical issues, but not the detailed procedures and paperwork associated with gaining approval from institutional review boards. Although HURAA is targeted for military personnel, it can be used by researchers in academia and government agencies because the principles of research ethics are widely applicable.

The design of HURAA was guided by a number of broader objectives. We sought to (1) organize information for researchers who need to understand the ethical issues associated with using human subjects, (2) deliver it over the Web in conformance with e-learning standards, and (3) enhance the information-delivery effectiveness with a good human–computer interface and intelligent tutoring system (ITS) modules. The first goal requires that the content of HURAA be precise and cover all aspects of human subjects protection. The second goal requires a client–server architecture that organizes the information in a hierarchical structure and that fits standards manageable by different learning management systems (LMS). The third goal requires a user interface that follows human–computer interface guidelines and an LMS that incorporates sophisticated pedagogical techniques.

This paper is divided into three parts. First, we describe the content of HURAA. Second, we describe the architecture of the server–client organization and implementation of ITS components. Finally, we report the results of studies that evaluate the effectiveness of HURAA.

This work was directly supported by contracts from the Office of Naval Research (N61339-01-C1006) and the Institute for Defense Analyses (AK-2-1801) and was partially supported by grants from the National Science Foundation (REC 0106965, REC 0126265, ITR 0325428) and the Department of Defense Multidisciplinary University Research Initiative (MURI) administered by ONR under Grant N00014-00-1-0600. Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of ONR, IDA, or NSF. Correspondence should be addressed to X. Hu., Psychology Department, 202 Psychology Building, University of Memphis, Memphis, TN 38152 (e-mail: xhu@memphis.edu).

CONTENT OF HURAA

The targeted users of HURAA are military personnel who conduct research with human participants. HURAA contains all related regulations from the Army, Navy, and Air Force. It organizes the Common Rule (NIH 45 CFR 46; 32 CFR 219), the Belmont Report's coverage of beneficence, justice, and respect for persons, and the "seven critical issues" that must be scrutinized when evaluating any case (Emanuel, Wendler, & Grady, 2000). HURAA has three types of content: original content, authored content, and *Point & Query* questions, and it delivers the content in different ways, depending on the nature of the content.

Original Content

HURAA allows access to original documents through natural language queries (Graesser, Hu, Person, Jackson, & Toth, in press). To make this possible, all documents are analyzed at the level of individual paragraphs and are indexed using three different methods: LSA vectors (Foltz, 1996; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998), glossary items, and inverse frequency weighted words. When a user types in a question in natural language, all three indexing methods are used to access the entire content and return high matching documents. The accuracy of the information retrieval and indexing schemes is reported in Graesser et al. (in press). Over 90% of the first-choice documents retrieved are perceived by users as relevant, and 50% are judged as informative.

Authored Content

In addition to the original documents, several modules were created for different purposes. They are *issues*, *lessons*, *historical overview*, *cases*, and *decision consequences*. The authored content is indexed in the same way as the original documents. It is also hierarchically organized in a format that can be accessed and used by different LMSs and ITS modules. For example, the *issues* module summarized all ethics guidelines into seven critical issues: social or scientific value, scientific validity, fair subject selection, favorable risk–benefit ratio, independent review, informed consent, and respect for enrolled subjects. These seven critical issues were further elaborated into 24 topics and 123 points.

Point & Query Questions

HURAA is a facility that users can interact with and that can provide them with helpful information. A context-sensitive question–answer facility, called *Point & Query* (P&Q), helps the user learn about the content more efficiently (Graesser, Langston, & Baggett, 1993). When the user clicks on a hot spot, a list of context-sensitive questions appears; the user selects a question and then receives an answer to that question. In the P&Q facility, HURAA organized 460 questions and answers, with a small set of two to six context-sensitive questions being tailored to each of the *topics* and *points* in the *issues* module.

Relations Between Contents

HURAA uses an LMS to present the three different types of content in an integrated, cross-classified manner. For example, when the user is exploring the *issues* module, each of the *critical issues*, *topics*, and *points* is further associated with different *cases*. Each of the *topics* and *points* has associated P&Q questions that are available to the users.

HURAA INTERFACE AND ARCHITECTURE

HURAA supports diverse methods of navigation and information retrieval. In addition to the typical features of Web sites (such as graphical user interfaces and hypertext), a number of unique intelligent features are now just starting to be used on Web sites, including an animated conversational agent, P&Q, and natural language queries. In this section, we first describe the major functions of HURAA interface and then present the client–server architecture of HURAA.

HURAA User Interface

The HURAA interface (see Figure 1) is similar to most Web pages, where visible space is separated into several functional areas. The left panel serves as the highest level of the table of contents, while the right panel occupies 80% of the visible space and serves as the main display area. The top left corner is reserved for the navigational agent. With such a layout, it is convenient to view the various forms of navigation and information retrieval from the perspective of the five approaches to navigation discussed by Quesenbery (2001). Table 1 identifies the different forms of navigation and retrieval in each of the modules and features of HURAA: (1) Browsing—"I want to see what's available"; (2) Finding—"I want to locate something specific"; (3) Query—"I want to see items that meet my criteria"; (4) Structured—"I want to be led through a series of choices to help me narrow my focus"; and (5) Guided—"I want to be led through the information." HURAA tries to implement all five of the features, as summarized in Table 1. For example, an animated agent guides the user during the interaction, whereas P&Q provides a structured method of allowing the user to query the system with a context-sensitive Frequently Asked Questions (FAQ) facility.

Animated conversational agent. This is a talking head with synthesized speech, facial expressions, and pointing gestures. The agent tells the user what to do next when he or she first encounters a Web page. For example, when the user enters the *explore issues* module, the agent says, "Select the issue that you would like to explore." The talking head moves to direct the user's attention to some point on the display and also tells the user what each module is supposed to do, after the mouse pointer rests over the module link for more than 2 sec.

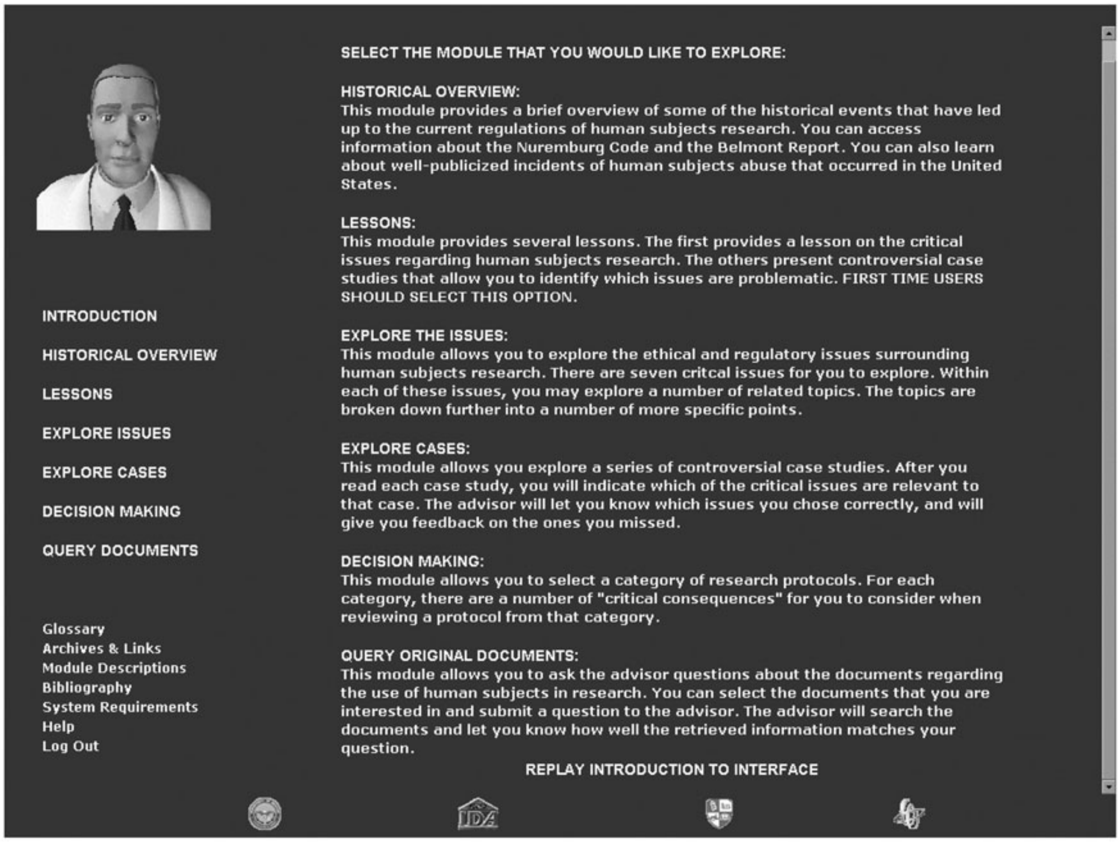

Figure 1. Home Page of HURAA.

The agent is designed to project an authoritative persona and to help the user navigate through the interface more quickly. Many novice users are lost and do not know what to do next when they encounter a page. The agent helps to reduce this wasted time in addition to enhancing motivation and possibly promoting deep learning (Atkinson, 2002; Baylor, 2001; Johnson, 2001; Moreno, Mayer, Spires, & Lester, 2002).

P&Q. The interface displays a question mark for some hot spots. When the user clicks on the question mark, a menu of questions appears. The user selects a question by clicking on it, and then receives the answer. Thus, the user can quickly ask a question with two clicks of a mouse. Users are also exposed to good, relevant questions because the designer of the P&Q carefully provides these question–answer items.

Table 1
Forms of Navigation and Retrieval for Different HURAA Modules and Features

Module/Feature	Browsing	Finding	Query	Structured	Guided
Introduction					×
Historical Overview	×	×	×	×	
Lessons			×	×	
Explore Issues	×	×	×	×	
Explore Cases	×	×		×	
Decision Making				×	
Query Documents			×		
Glossary	×	×	×		
Archives & Links	×	×			
Module Description	×				
Bibliography	×				
System Requirements	×				
Help	×				
Point & Query	×	×	×	×	

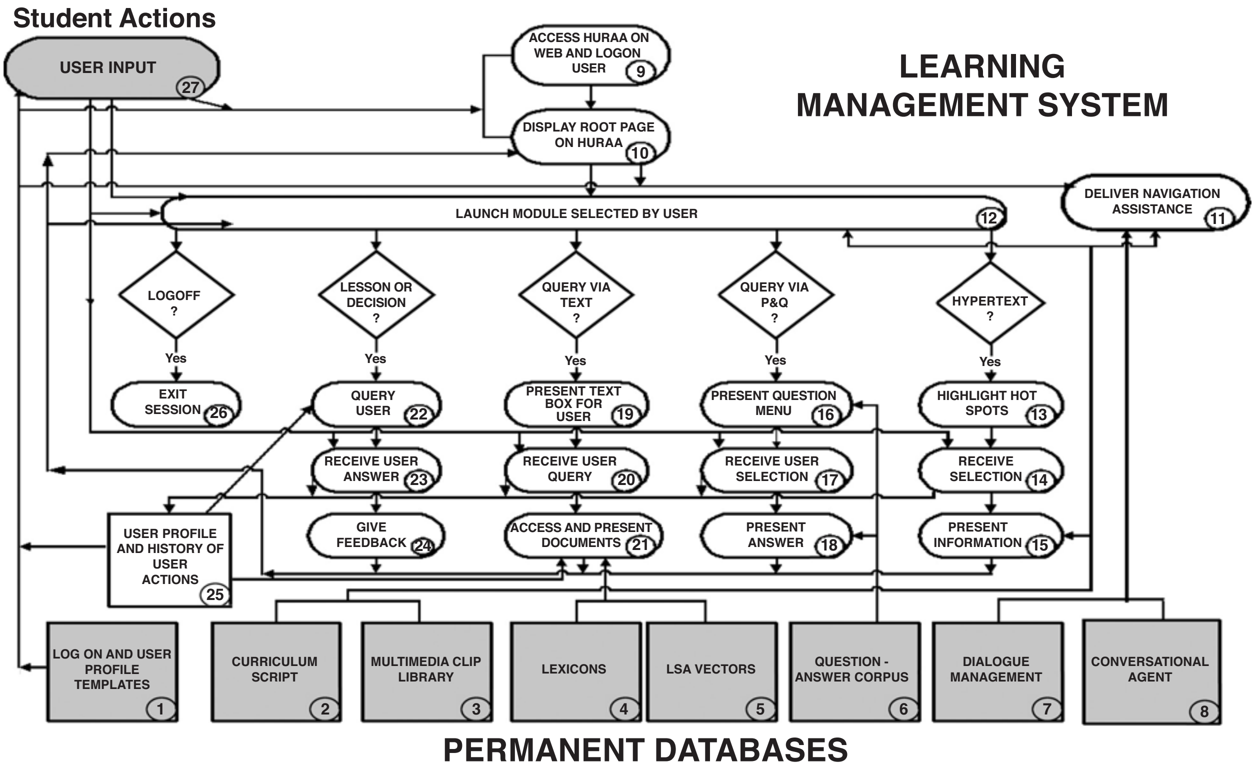

Figure 2. Architecture of HURAA.

HURAA Architecture

Figure 2 presents the architecture of HURAA in diagram form. At the bottom (Components 1–8) are the permanent databases that are not modified while the user interacts with the Web site. These databases can, of course, be modified as needed by personnel maintaining the HURAA system. HURAA's actions (processes and computational procedures) are depicted in Components 9–26, which constitute the Learning Management System of HURAA. Component 25 is a dynamic storage unit that includes the user's profile and his or her history of prior use of the Web facility. The dynamic storage registers are continuously updated as the HURAA–user interaction proceeds. Component 27 refers to the user's input, such as mouse clicks and English messages that are typed in.

A few words should be said about the notation in Figure 2. The components in rectangles represent permanent databases or dynamic storage units, whereas those in the diamonds depict decisions and those in the rounded boxes depict processes. The arrows that connect actions depict the temporal flow of processes. An arrow from a permanent database (D) or a dynamic storage unit (DSU) to an action (A) means that the contents of {D|DSU} are accessed or recruited to execute action A. An arrow from A to a DSU means that some of the results of A's computations are stored in the DSU. The permanent databases should be viewed as "read only," so the results of actions are never sent to the permanent databases.

Components of HURAA. This subsection provides a technical description of the 27 components of HURAA. It starts with the permanent databases and then turns to the LMS. This architecture and the HURAA components were designed to conform with the DoD Advanced Distributed Learning (ADL) initiative and its specification for reusable instructional objects, the Sharable Content Object Reference Model (SCORM, Versions 1.1, 1.2, and 1.3, 2001). Conversion programs were used to develop permanent databases for HURAA. These programs create XML code from HTML code that has been augmented by metalevel descriptions associated with SCORM. At present, the entire set of capabilities of HURAA requires the user to have Microsoft Internet Explorer (IE) 5.5 and a high-speed connection. The URL of HURAA is http://www.huraa.net.

Permanent databases. The eight permanent databases are presented in the bottom of Figure 2. These databases and other HURAA components in Figure 2 have been assigned numbers that are displayed in the lower right-hand corner of each component box or oval. Among the eight permanent databases, some (such as 1, 2, 3, and 6) are similar to most of the database-driven, client–server Web applications. Some of the modules are specially designed for HURAA implementation. What follows is a brief description of module 4 (lexicons) and module 5 (LSA vectors).

The *lexicon* module is a corpus of words that have particular features associated with each lexical entry. The most important lexicon in the current version of HURAA consists of the word-frequency norms of Francis and Kučera (1982), which are available electronically. Each word entry includes its part of speech (noun, adjective, main verb, preposition, etc.) and its frequency of usage in the English language. The lexicon is used in the information retrieval algorithms of Component 21. The lexicon is stored either in flat ASCII text files or in a structured data file that can be accessed through Microsoft's SQL server.

The *LSA Vectors* module is created for each word and document. LSA has recently been proposed as a statistical representation of a large body of world knowledge (Foltz, 1996; Landauer & Dumais, 1997; Landauer et al., 1998). An LSA space can be created and tested for a very large corpus of documents in a short period of time (measured in days). From the present standpoint, LSA was used as an information retrieval mechanism for accessing paragraphs in the document space that are relevant to user questions (either in English or by the P&Q facility). It is beyond the scope of this article to describe LSA, in part because this technology has been discussed extensively elsewhere (Graesser et al., in press; Landauer et al., 1998). It suffices to say that HURAA used LSA as a method of representing world knowledge in the information retrieval component that accommodates questions in natural language. The document retrieval algorithm identified a particular document in a large corpus of information sources that was highly relevant to a user's query *Q*. Document paragraph *P* was computed as the best match with query *Q,* with respect to the cosine match between their LSA vectors, max {cosine(*P, Q*)}. An LSA vector was stored for each word in the lexicon (component 4) and for each paragraph in most of the modules of the Curriculum Script (Component 2). The values of the LSA vectors are real numbers, and they took up the vast majority of the storage space in HURAA, approximately 400 MB or 67% of the facility.

LMS. The modules in Figure 2 are managed by an LMS server application. The LMS has one dynamic storage unit (Component 25) and 17 procedures (actions or processes) that are executed (Components 9–24, 26). HURAA operates similarly to most Web-based resources in which an LMS manages a user profile, records a learning history, and selects appropriate contents to deliver to the user. The LMS in HURAA adds a few ITS enhancements which make HURAA different from most Web facilities. Some of these ITS enhancements are summarized below.

The *issues and cases matrix* supports case-based reasoning. There are two authored modules that have a similar structure but serve two different purposes. The *issues* module organizes ethical principles in terms of a tree structure of *critical issues*, *topics*, and *points.* The *cases* module is organized along the same lines, but it is composed entirely of examples that have links to the *issues*, *topics*, or *points*. For example, when the user is reading a topic about informed consent, a link will appear for a case in which the informed consent is violated.

The *natural language query* module was described earlier. The three different types of content (original documents, authored content, and authoring questions) are indexed at the level of each paragraph. Such indexing makes it possible for the LMS to obtain appropriate content any time the user poses a question in English. HURAA accesses five paragraphs from the source documents that should contain information relevant to the query.

A *dynamic selection of lessons* provides instruction that adapts to the learner. This user-modeling ITS enhancement is implemented in the *lessons* module. In the lessons module, a case (e.g., the *Tuskegee Syphilis Study*) is first presented to the user, and then the user is prompted to review the seven major issues. After that, a "test your knowledge" page is presented to the user with seven choices, corresponding to each of the seven critical issues. The user must decide whether a given critical issue is problematic in the particular case. The input is then compared with expert answers regarding that specific issue. The LMS provides feedback to the user and records the answers, which are used to select the next appropriate lesson for the user. This feature is implemented through the use of a learning profile and dialogue history (Component 25 of Figure 2).

EMPIRICAL TESTS OF HURAA

We have evaluated many of the features of HURAA in empirical studies. Highlights of three of these studies are summarized in this section. One study by Graesser et al. (2002) compared HURAA with a standard Web facility on assessment of memory for the material, case-based reasoning, searching for answers to questions, and impressions of the learning experience. This section reports new data with a larger sample of participants than were reported in Graesser et al. (2002). Graesser, Ventura, et al. (2003) evaluated the impact of the animated conversational agent of HURAA on the same outcome measures by comparing it with navigational guides in alternative media (e.g., voice only or printed text). Graesser et al. (in press) evaluated the quality of HURAA's answers to user questions posed in natural language.

How does HURAA Compare with Conventional Computer-based Training?

Graesser et al. (2002) compared HURAA with the same content presented in a conventional computer-based training environment. The participants were 18 graduate students and upper-division undergraduates at the University of Memphis and Rhodes College who were seeking approval of their own research projects from a local IRB. Consequently, they had a vested interest in understanding the ethics of human experimentation. The participants received introductory material on ethics and HURAA, completed lessons that involved case-based and explanation-based reasoning, and performed searches for answers to test questions. Nine participants received the full HURAA version and 9 a Standard web version with conventional page-turning and hypertext capabilities. The Standard version was designed to have information equivalence to the HURAA version so that any differences in performance could not be attributed to an imbalance of information available to the learner. The data reported in the present study was on an expanded set of 26 participants, 13 in each of the two conditions.

The HURAA and Standard versions were compared on 13 measures that were collected either during or after training. Means, standard deviations, and effect sizes for these measures are presented in Table 2. The nature of these tests is briefly described here, whereas more details can be found in Graesser et al. (2002).

Memory for introductory material. The material in the introduction of HURAA was a professionally constructed multimedia presentation, whereas the same verbal content was presented in printed text in the Standard version. Memory was assessed in three tests: free recall, cued recall, and the cloze task. The free recall presented a series of concepts that the participants were asked to define or describe in writing in a booklet. For example, they were asked "What are the three principles of the Belmont Report?" After the free recall task, the cued recall test was administered. It had more retrieval cues than the free recall test did—for example, "Autonomy is one of the three principles of the Belmont Report. Describe Autonomy." The cloze procedure had the most retrieval cues. Verbatim segments of the introductory text were presented, but with some critical words left blank; the participant wrote down the words in the blanks. There were progressively more retrieval cues for content to be retrieved as one went from free recall to cued recall to the cloze task.

As can be seen in Table 2, memory was found to be better for the HURAA version than for the Standard version. The participants assigned to the HURAA condition had significantly higher free recall proportions [$F(1,24) = 12.87$, $MS_e = .02$, $p < .05$] and higher cued recall proportions [$F(1,24) = 2.88$, $MS_e = .04$, $p < .05$, one-tailed]; the cloze recall proportions were in the predicted direction but not quite significant [$F(1,24) = 2.03$, $MS_e = .02$, $.05 < p < .10$, one tailed]. The multimedia effects of HURAA, which were designed to be captivating and persuasive, apparently did have a significant impact on memory for the introductory material. It should be noted that there was no significant difference in the time spent comprehending the introduction in the two versions, so time on task could not explain the recall differences.

Identification of potentially problematic issues in cases. As discussed earlier, there are seven critical issues that must be considered whenever the experiments are to be evaluated on research ethics (e.g., social or scientific value, fair subject selection, etc.). The posttest assessed how discriminating the participants were in identifying potentially problematic issues in two cases. The cases were selected systematically so that six of the issues were problematic in only one of the two cases; one of the issues was problematic in both cases, so it was not scored. This test was functionally a transfer test from the case-based, explanation-based reasoning task in the *lesson* module of the acquisition phase. During this transfer test, the participants simply read each case and rated the seven issues on a 6-point scale (as to whether issue I

Table 2
Means and Standard Deviations for Dependent Measures in Experiment 2

	Condition				
	Full HURAA (n = 13)		Standard Web (n = 13)		
Dependent Measures	*M*	*SD*	*M*	*SD*	Effect Size
Memory for core concepts					
Free recall proportion	.66	.10	.47	.16	1.19
Cued recall proportion	.64	.14	.50	.25	.56
Cloze recall proportion	.50	.17	.39	.19	.58
Problematic issue identification					
Hit proportion	.64	.10	.64	.10	
False alarm proportion	.34	.15	.30	.16	
d' score (discrimination)	.83	.65	.71	.85	.14
Task completion time (minutes)	21.2	4.0	19.6	5.8	
Search for information					
Correct document retrieval proportion	.76	.18	.60	.24	.67
Search time (minutes)	21.5	8.6	21.0	5.8	−.09
User impression ratings					
Amount learned	4.54	1.05	4.54	1.27	.00
Interest	4.54	1.51	4.54	1.05	.00
Enjoyment	3.62	1.39	3.85	1.14	−.20
Ease of learning	3.92	1.04	3.92	1.12	.00

*n = number of participants.

was problematic for case C): 1 = *definitely not a problem*, 2 = *most likely not a problem*, 3 = *undecided, guess it's not a problem*, 4 = *undecided, guess it's a problem*, 5 = *most likely a problem*, and 6 = *definitely a problem*. Hit rates, false alarm rates, and d' scores were collected for each participant and analyzed. The gold standard for deciding whether an issue was truly problematic for a case was based on the decisions of seven experts on research ethics in the Department of Defense.

The results uncovered no statistical differences between the HURAA and Standard versions. The hit rates, false alarm rates, and d' scores were nearly identical in the two conditions (all Fs < 1). Task completion times did not significantly differ between the two conditions on the issue identification task ($F < 1$). It appears that there was no pedagogical advantage in having participants actively decide which issues were problematic in particular cases and receive explanations about discrepancies with experts (as implemented in the HURAA version during training). It was just as effective to have the participants simply read the experts' decisions and explanations for each of the seven issues. It should be noted, however, that there were only four cases in the *lesson* module that involved case-based reasoning during training. Differences might emerge if the *lesson* module of HURAA had a dozen or more cases.

Retrieval of answers to questions. We measured the speed and accuracy of retrieving answers to the four questions given to the participants during the training phase. For example, two of the test questions were "Find out what the specific limitations are on the uses of prisoners in research" and "Find out what a medical monitor is in human subjects research." The amount of time it took to find answers to the four questions did not significantly differ for the HURAA and Standard versions. However, the likelihood of retrieving a document with the correct answer was significantly higher in the HURAA condition than in the Standard condition [$F(1,24) = 4.02$, $MS_e = .04$, $p < .05$]. Therefore, the natural language queries posed by participants in HURAA produced a much higher accuracy for fetching relevant documents than did the conventional methods of searching documents for answers in computer-based environments.

User impression ratings. At the end of the experiment, the participants were asked to rate the learning environment of four 6-point scales: amount learned, interest, enjoyment, and ease of learning. The high numbers indicated more positive impressions. These ratings leaned to the positive side (i.e., above 3.5, the center), but there were no significant differences between the HURAA and Standard versions.

In summary, the HURAA version was superior in improving performance in the memory task and the document search task, whereas case-based reasoning and user impressions were unaffected. The effect sizes on the impact of HURAA on memory and successful document retrieval varied from .56 to 1.19, whereas the effect sizes hovered around zero for problematic issue identification, user impressions, and task completion times.

How Important Is the Animated Conversational Agent as a Navigational Guide?

Graesser, Ventura et al. (2003) evaluated the impact of the animated conversational agent of HURAA on the same outcome measures reported in Table 2. Participants were randomly assigned to one of the following four conditions: full HURAA guide, voice guide, print guide, and no guide. The participants were 155 university undergraduates who participated for extra credit in a course or for $20. If a navigational guide is important, then completion of the various tasks should be poorer in the no-guide condition than in the other three conditions. If speech is a particularly important feature among the three conditions with a navigational guide, then the prediction would be print < min{HURAA, voice}; if print is superior, then the prediction would be the opposite. If the presence of the face improves interactivity, then the prediction is HURAA > voice; if the face is a distraction from the material in the main display, then the prediction is the opposite.

The most striking results of Graesser, Ventura et al.'s (2003) study were the lack of significant differences among conditions. In fact, there were no significant differences among the conditions for any of the 13 dependent measures. Thus, the use of agents as navigational guides on a Web site has not yet proven to be effective. However, there is some evidence that these agents have a positive impact on learning or on one's perceptions of the learning experience when subject matter is presented with the agents, rather than as speech alone or as printed text controls (Atkinson, 2002; Baylor, 2001; Graesser, Moreno, et al., 2003; Moreno et al., 2002; Whittaker, 2003). The presentation of learning content may sometimes be facilitated by agents, whereas agents may not be so effective as navigational guides.

How Good Are HURAA's Answers to Users' Questions in the Natural Language Retrieval Facility?

One goal of HURAA was to optimize information retrieval when users actively attempted to search for information. Some users prefer to learn by actively exploring information, asking questions, and accessing answers to questions (Graesser & Person, 1994; Otero & Graesser, 2001), whereas others are more passive and rarely pose questions. A good Web facility provides information quickly on demand for those occasions when users are curious, active, self-regulated learners.

HURAA implemented multiple methods of accessing information in a large space of documents. HURAA has hypertext and glossaries, which are conventional methods of information retrieval. It has a *Point & Query* facility (Graesser et al., 1993), answers to context-sensitive Frequently Asked Questions (FAQs). HURAA also accesses documents that provide answers to questions that the learner formulates in English, using techniques in computational linguistics (Jurafsky & Martin, 2000; Landauer et al., 1998). It is the latter facility (i.e., document access via queries in natural language) that motivated the present analyses. We evaluated the performance of our natural language document retrieval facility in supplying helpful and relevant information.

The natural language query feature is the most direct form of information retrieval because the user does not need to depend on or wait for the Web content and interface to set the stage for posing a particular question. The user simply asks the question in English, at any point in the session with HURAA. In contrast, all of the other information retrieval facilities require the user to sift through the Web pages, menu options, or hypertext hot spots until there is a screen where a particular question can be posed, if indeed it can be posed at all. Because of this special status of the natural language query feature, Graesser et al. (in press) conducted a performance analysis on how well different retrieval algorithms fetched relevant and helpful information.

Graesser et al. (in press) performed a series of experiments that tested eight different retrieval algorithms. These alternative algorithms consist of all combinations of the presence versus absence of three match functions: (1) LSA, (2) inverse frequency weighted word matching (called *Word*), and (3) glossary word matches (called *Glossary*). Each match function produced a value of 0 to 1 when considering the overlap between the words in query Q and the words in paragraph P. The paragraph with the highest overlap, i.e., max $\{\text{overlap}(Q, P)\}$, was returned when each one of the three functions was considered. When multiple match functions were considered (such as LSA + Glossary), then the paragraph was selected that had the highest average match values from the multiple functions. The study also considered context in these matches. The context included the dialogue history (the words affiliated with the last three actions of the user) and/or the Web page information (the words affiliated with the current Web page). The words in the context information were added to the words in query Q when matches were computed. This context information in the computation might improve the relevance and informativity of the highest matching paragraph P.

Experts on research ethics were instructed to generate questions about HURAA periodically while using the Web facility. The experts were instructed to generate any question that they wished, as long as it was not a yes/no question or a disjunctive question (e.g., "Is A or B the case?"). The experts were told that open-ended questions that invite lengthy answers were best. The experts entered each question, one at a time, and HURAA supplied a sample of answers. Each answer consisted of a high matching paragraph document from the designated document space and was rated by the experts on the two 6-point scales. The *relevance rating* had the following 6 points: 1 = *very irrelevant*, 2 = *somewhat irrelevant*, 3 = *undecided but guess it's irrelevant*, 4 = *undecided, but guess it's relevant*, 5 = *somewhat relevant*, and 6 = *very relevant*. A *relevance proportion* was computed over a set of answers, which consisted of the proportion of relevance ratings that were 4, 5, or 6. Similarly, the *informativity rating* had an analogous set of 6 points, and an *informativity proportion* was computed over a set of answers.

The results of the analyses of three experiments revealed that the Word and LSA algorithms yielded approximately the same performance with respect to the ratings, and that the Glossary algorithm was substantially inferior. For example, in one experiment, the mean relevance proportions were .78, .69, and .70 for the LSA + Word, LSA, and Word algorithms, respectively; the corresponding mean informativity proportions were .42, .41, and .38. The ratings followed trends similar to those for the proportions. The Glossary algorithm was comparatively low in another experiment, where relevance proportions were .66, .22, and .06 for Word, Glossary, and a random selection of paragraphs, respectively.

What was particularly robust in the results was the profound effect of the context of the words on the screen that launched the question (called *screen context*). The means were much higher when the words in the screen context were present versus absent in the match algorithm; the means were quite different in the case of relevance proportions (.90 vs. .55), and informativity proportions (.50 vs. .31). In contrast, the dialogue history was not a particularly important form of context.

The robust impact of screen context on the success of document retrieval has two important practical implications. The first is that commercial information retrieval facilities are losing an important source of information when they ignore the context that launches searches for information. Second, the retrieval algorithms are impressive when there is a question context. Relevance proportions of .90 and informativity proportions of .50 are indeed impressive! The small differences in the match functions (LSA vs. Word) suggest that it sometimes is not worthwhile to go to the expense and effort of using LSA. Overall, the inverse frequency weighted word algorithm alone was about 90% as good as the hybrid algorithm that combined LSA and Word, whereas LSA alone was not better than Word alone.

CLOSING COMMENTS

The design of HURAA was inspired by multiple goals and needs of users. As a consequence, it ended up being a complex help facility that motivated the learner, trained the learner on core concepts, supported information retrieval and inquiry, and served as a large repository of information. HURAA is not a simple Web facility that satisfies a single, specific pedagogical theory. We believe that HURAA is an exemplar of future learning environments that will be used in the real world.

Colleagues have inquired how difficult it would be to develop a similar system for subject matter other than research ethics, the focus of HURAA. Our estimate is that the architecture is sufficiently generic that a learning environment on a new topic could be developed in 1–3 months. A new repository of documents would need to be created, so a corpus of electronic documents would save time. The word-matching algorithm for information retrieval from natural language queries would remain unchanged because the word frequencies are extracted from a generic word frequency lexicon (Francis & Kučera, 1982). The content of all other modules are created with the assistance of an EBOOK authoring tool (Hu, Mathews, Graesser, & Susarla, 2002), a tool that requires approximately a week of training and practice to master. Therefore, as with all learning systems, the primary bottleneck in development time lies in the process of a subject matter expert proficiently using authoring tools. The animated conversational agent accepts any arbitrary text input that is supplied by the authoring tool, so this component, unlike recorded speech, involves no new development time.

REFERENCES

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology,* **94**, 416-427.

Baylor, A. L. (2001, May). *Investigating multiple pedagogical perspectives through MIMIC (Multiple Intelligent Mentors Instructing Collaboratively)*. Workshop on agents at the AI-ED International Conference, San Antonio, TX.

32 CFR 219 (1991). *Protection of human subjects*. Department of Defense. Retrieved June 30, 2004 from DoD Web site http://www.access.gpo.gov/mara/cfr/waisidix_99/32cfr219_99.html/.

DoD Directive 3216.2 (1993). *Protection of human subjects in DoD supported research*. Department of Defense. Retrieved June 30, 2004 from DoD Web site http://www.dtic.mil/whs/directives/corres/html2/d32162x.htm/.

Emanuel, E. J., Wendler, D., & Grady, C. (2000). What makes clinical research ethical? *Journal of the American Medical Association,* **283**, 2701-2711.

Foltz, P. W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, & Computers,* **28**, 197-202.

Francis, W. N., & Kučera, N. (1982). *Frequency analysis of English usage*. Boston: Houghton-Mifflin.

Graesser, A. C., Hu, X., Person, P., Jackson, T., & Toth, J. (in press). Modules and information retrieval facilities of the Human Use Regulatory Affairs Advisor (HURAA). *International Journal on eLearning*.

Graesser, A. C., Hu, X., Person, N. K., Stewart, C., Toth, J., Jackson, G. T., Susarla, S., & Ventura, M. (2002). Learning about the ethical treatment of human subjects in experiments on a Web facility with a conversational agent and ITS components. In S. A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), *Intelligent Tutoring Systems 2002* (pp. 972-981). Berlin: Springer-Verlag.

Graesser, A. C., Langston, M. C., & Baggett, W. B. (1993). Exploring information about concepts by asking questions. In G. V. Nakamura, R. M. Taraban, & D. Medin (Eds.), *The psychology of learning and motivation: Vol. 29. Categorization by humans and machines* (pp. 411-436). San Diego: Academic Press.

Graesser, A.C., Moreno, K., Marineau, J., Adcock, A., Olney, A., & Person, N. (2003). AutoTutor improves deep learning of computer literacy: Is it the dialog or the talking head? In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Proceedings of Artificial Intelligence in Education* (pp. 47-54). Amsterdam: IOS Press.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal,* **31**, 104-137.

Graesser, A. C., Ventura, M., Jackson, G. T., Mueller, J., Hu, X., & Person, N. [K.] (2003). The impact of conversational navigational guides on the learning, use, and perceptions of users of a Web site. *Proceedings of the AAAI Spring Symposium 2003 on Agent-Mediated Knowledge Management*. Menlo Park, CA: AAAI Press.

Hu, X., Mathews, E., Graesser, A. C., & Susarla, S. (2002). EBOOK.EXE: A desktop authoring tool for HURAA. M. Reeves & T. C. Reeves (Eds.), *Proceedings for E-Learning 2002: World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education* (pp. 471-476). Montreal.

Johnson, W. L. (2001). Pedagogical agent research at CARTE. *AI Magazine,* **22**, 85-94.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice-Hall.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review,* **104**, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes,* **25**, 259-284.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2002). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition & Instruction,* **19**, 177-213.

NIH 45 CFR 46 (1991). *National Institutes of Health Code of Federal Regulations, Protection of Human Subjects*. Retrieved June 30, 2004 from NIH Web site http://ohrp.osophs.dhhs.gov/humansubjects/guidance/45cfr46.htm/.

Otero, J., & Graesser, A. C. (2001). PREG: Elements of a model of question asking. *Cognition & Instruction,* **19**, 143-175.

Quesenbery, W. (2001). On beyond help: Meeting user needs for useful online information. *Technical Communication,* **48**, 182-188.

Sharable Content Object Reference Model (SCORM), Version 1.1. www.adlnet.org.

Whittaker, S. (2003). Mediated communication. In A. C. Graesser, M. A. Gernsbacher, and S. A. Goldman (Eds.), *Handbook of discourse processes* (pp. 243-286). Hillsdale, NJ: Erlbaum.

(Manuscript received December 12, 2003;
revision accepted for publication April 6, 2004.)