RC23941 (W0604-123) April 25, 2006
Computer Science

IBM Research Report

An Experimental Investigation of the Effectiveness of Individualized Web-based Learning Based on the Dynamic Assembly of Learning Objects

John C. Thomas, Robert Farrell
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598

An Experimental Investigation of the Effectiveness of Individualized Web-based Learning based on the Dynamic Assembly of Learning Objects

Authors: John C. Thomas, Robert Farrell

IBM T. J. Watson Research Center

P.O. Box 704

Yorktown Heights, New York 10598

jcthomas@us.ibm.com, robfarr@us.ibm.com

September 3, 2004

Technical Lead: Robert Farrell

Other Team Members have included: Sam Dooley, Bill Rubin, Stephen Levy, Amy Katriel, Doug Gordin, and Danny Oppenheim

Table of Contents

Abstract

This research presents an experimental evaluation of a particular approach to the problem of organizing information for effective learning. The approach divides the task of organizing information between the user and the system, leveraging expert judgments of how information should be modularized and recombined. We have implemented the approach in a web-based self-paced learning system called Custom Course. An experiment was conducted to see whether the automatic information organization and user interaction techniques employed in the system could result in improved outcomes compared to a typical text information retrieval system. Users of similar experience levels performed significantly better on a design task after using the Custom Course system, according to independent expert judgments of design products. The results suggest that actively engaging users in building a learning path that is also informed by expert judgments provides an effective personalized structure for learners. Reading hypermedia in the form of well-organized personalized learning paths may improve comprehension by integrating new, complex information into the user's existing mental schemas. The approach could be useful for improving performance on tasks requiring time-constrained comprehension.

Introduction

With the explosive growth of electronically stored materials, both those that are explicitly educational and those that could be used for educational purposes (though originally constructed for other reasons), the need arises for selecting and organize specific subsets of materials appropriate to a particular learner's goals, background and abilities. Such assemblages could be generated from a wide variety of materials and made available to a learner quickly via the web. In order to achieve such a vision, three major sets of issues must be addressed. First, there are intellectual property issues. Digital rights need to be associated with materials as they are (re)assembled and usage needs to be regulated and enforced in different contexts. In this paper we will not address these issues. Second, in order for material to be selected appropriately, it will be necessary to break large materials into smaller chunks ("modules") and to add descriptive metadata to these chunks to create "Learning Objects". Significant progress has recently been made on agreement about the metadata. The Learning Object Metadata (LOM) is an IEEE standard that provides a powerful description of objects related to training and education. Third, some mechanism must be designed and implemented to select appropriate modules from the potentially vast number of modules and then organize these modules into some pedagogically meaningful sequence. Ideally, this process will be guided by the learner's needs and capabilities but without requiring a large, up-front investment of time and effort on the part of the learner to communicate these needs and capabilities to the system.

Project Description

This project primarily addresses the issues of Learning Object representation, selection, and sequencing. For the purposes of this project, we are working with internally published materials for which the owner has the copyright and right to reassemble and recombine the content in any way and is not concerned about payment, thus reducing the need for complex digital rights management.

To address issue of Learning Object represetation, we have been primarily using XML standards for digital content and extending the open IEEE Learning Object Metadata standard. In addition to the question of what metadata is needed, there is also the issue that adding metadata takes work. There are at least five basic approaches to adding metadata. First, one could conceive of tools that would help creators of original materials add metadata at the time of creation to enable broader use. This might be particularly useful if adding the metadata is seen to improve the quality of the process or product and not as a separate process requiring "extra" work. A second possibility is that experts could break materials into smaller chunks and add metadata as a separate step. A third possibility is that metadata might be calculated algorithmically from the materials themselves, using techniques such as extracting and transforming the source dataformats, data mining, and natural language processing. A fourth possibility is that metadata may be assigned dynamically and algorithmically by the learning environment or system during usage. The fifth possbily is that metadata might be entered manually by learners as learning objects are used or perhaps shared in a collaborative learning setting. In our project, we have used a combination of all of these techniques.

The issues surrounding the effective selection and sequencing of educational materials are complex. Expert human tutors preparing materials for a specific learner with specific knowledge, background, and goals might bring to bear almost any life experience as well as educational principles in performing this task. While there are clearly known benefits to human tutoring (Bloom, 1984), most people do not have unlimited access to excellent human tutors or intelligent tutoring systems specifically tuned for a specific domain and set of tasks (Farrell, Anderson, and Reiser, 1985; Anderson, Boyle, Farrell, and Resier, 1987). Instead, instruction is often done in classroom settings that may be far from ideal for any specific individual. Alternatively, the adult motivated learner may not have access to *any* timely classes focused on their specific needs and must learn what they can from books, on-line materials, or other people. In each case, finding the information that is at the right level and appropriate to current needs can be challenging despite the availability of search engines.

In the case of our project, we decided to focus initially on adult, motivated learners with some information technology (IT) background who have specific educational needs that are job-related. This focus is important because it leads to learners who are motivated to find, organize, and comprehend. The idea that learners could perform some of the work to self-organize their learning is an integral part of our philosophy: have the system (or tutoring agent) and the learner work coopertively and interactively to select and organize materials to meet the learner's needs. The materials (taken from IBM technical books and presentations) dealt with the IBM WebSphere product and web services technologies, two important areas within IBM and of interest in the IT industry. We speculate that IT materials are mostly "objective" in content and have a logical progression. Our implementation could be extended to deal with areas such as sales tactics or business decision-making. However, the algorithms for assembling the materials would need to be altered considerably to address subjective and multi-perspective domains, such as literary criticism or history or to deal with other sets of users such as novices or very young children. By focusing on IT tutorials, we have identified a useful genre of instruction that utilizes materials that change quickly in response to new trends and developments in technology.

Software Description

The Custom Course system supports assembly, archiving, and sharing of personalized sequences of learning materials (see Figure 1). At the time of the experiment the basic sequence of events leading to a custom course was fairly straightforward. First, users entered keywords into a query field on the Course Assembly page (see Figure 2). This is an interface familiar to our target audience. Given the keywords, the search engine working over the full text of the modules in the repository returned ten results in order of relevance to the query; each result included a title, a summary description, an expected duration, and a difficulty level (see Figure 3). The learner then selected a subset of these results as being most relevant to their current learning needs. As the search result items were selected for inclusion in the learner's custom course, the total duration of the course (e.g., "Course Duration: 61 minutes")was displayed. Once the learner's final selections were made, the software arranged the selected modules in a new, suggested progression.

At the time of the experiment, the only information used by the system to sequence the results into the progression was the material's original order in the source materials and a logical learning sequence provided by experts -- an ordering of the books in a logical learning progression for the average target learner. Thus, the system was essentially taking segments of books and putting them back into a preferred reading order combining the judgement of both the author of the book (presumably an expert in the material) and an instructional design expert who defines a logical learning progression from basic to more advanced material. (In later versions of the system, we have experimented with sequencing strategies for search results based upon a wide range of learning object metadata, including topic, duration and resource type as well as our own list of rhetorical constructs, cognitive levels of processing, and IT industry categories such as code listings and architecture.)

The system used the sequence of search results to form the "custom course" (see Figure 4) consisting of a sequence of numbered lessons including one or two learning objectives listed for each lesson. A typical custom course thus consisted of between one and ten modules arranged as "lessons" in a progression, each lesson consisting of the task of reading particular learning material extracted from a book. In keeping with the general philosophy of having the learner in control, the order of lessons was only a suggested order. Learners were free to go through the lessons in any order and to skip lessons that seemed irrelevant.

Once the learner selected a specific lesson to view, the beginning of that section of material appeared on the screen (see Figure 5). At the top of the page was a clickable reference to the original source. Many lessons also had additional hyperlinks. Again, the learners were free to follow these links if they felt that was more productive to their purposes.

The current system (shown in Figures 1 through 5) has been renamed Dynamic Learning Experience (DLE) and has been enhanced in several ways over the system that was available to learners during the experiment. (For a more recent and more technical view of the current system, see Farrell, Liburd, and Thomas, 2004). In particular, in Figure 1, during the experiment, there was no Offered Courses hyperlink. In Figure 2, during the experiment, there were no advanced query options and there was no "Automatic

Assembly" tab. In Figures 3 and 4, modules did not have the extra information in brackets next to each item, namely an instructional use (e.g., "Introduction") and a topic (e.g.,"Java 2"). In Figure 5, the objective was not displayed with the learning material.

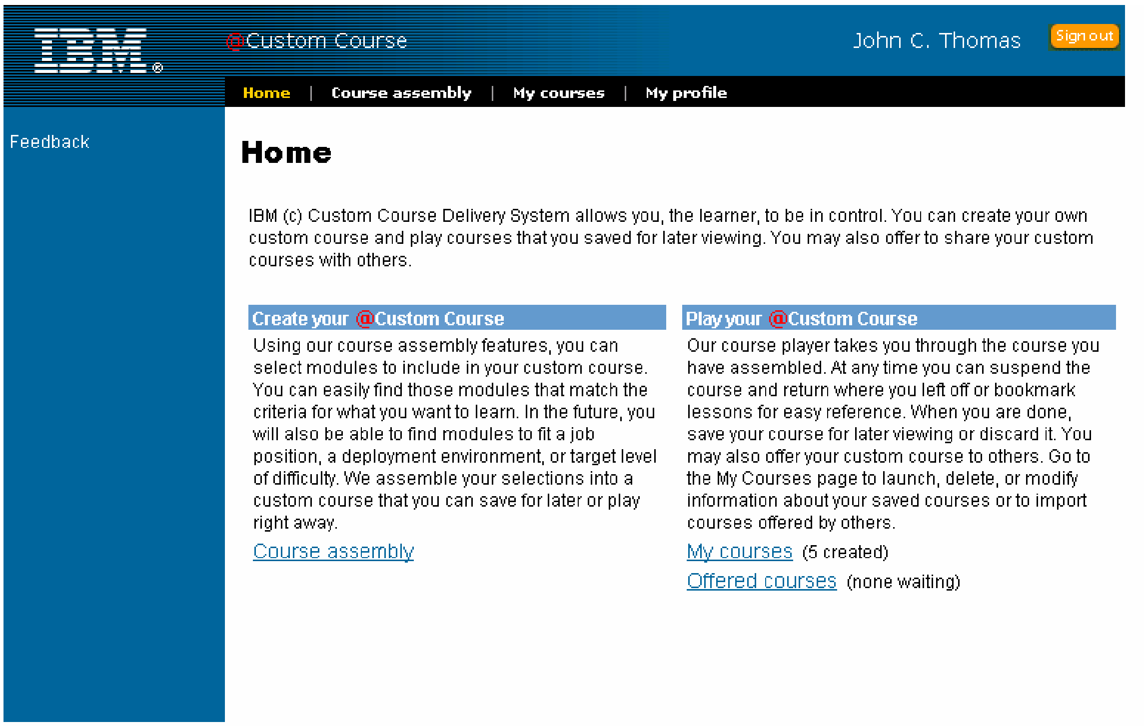

Figure 1: Home Page of the Custom Course software system

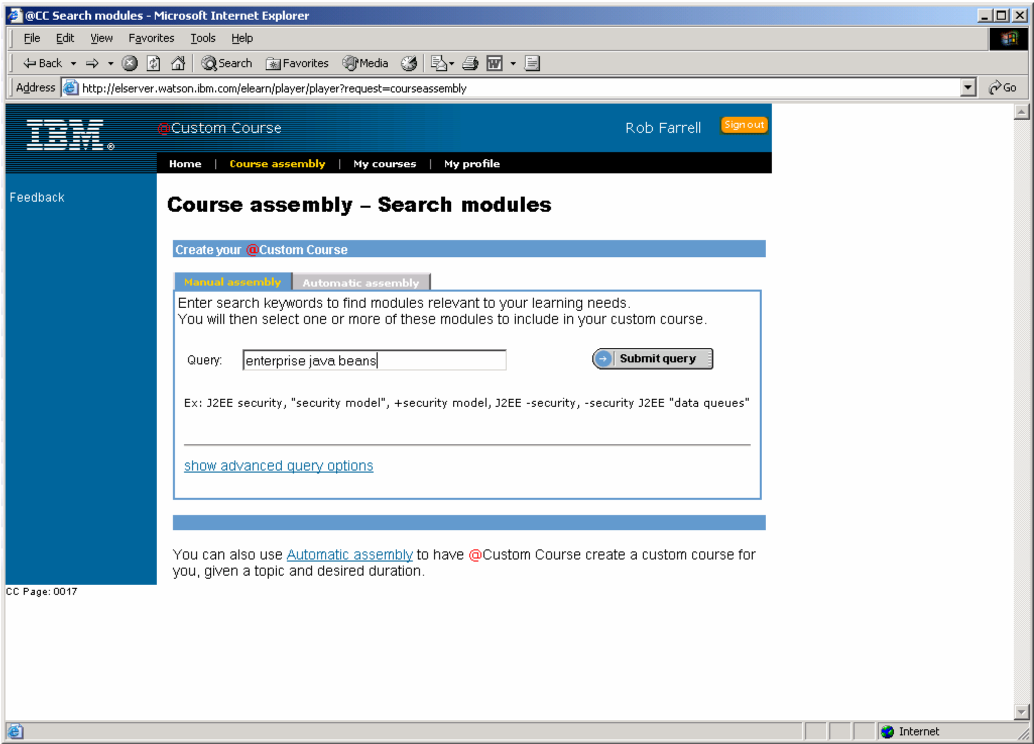

Figure 2: Query Page of the Custom Course software system

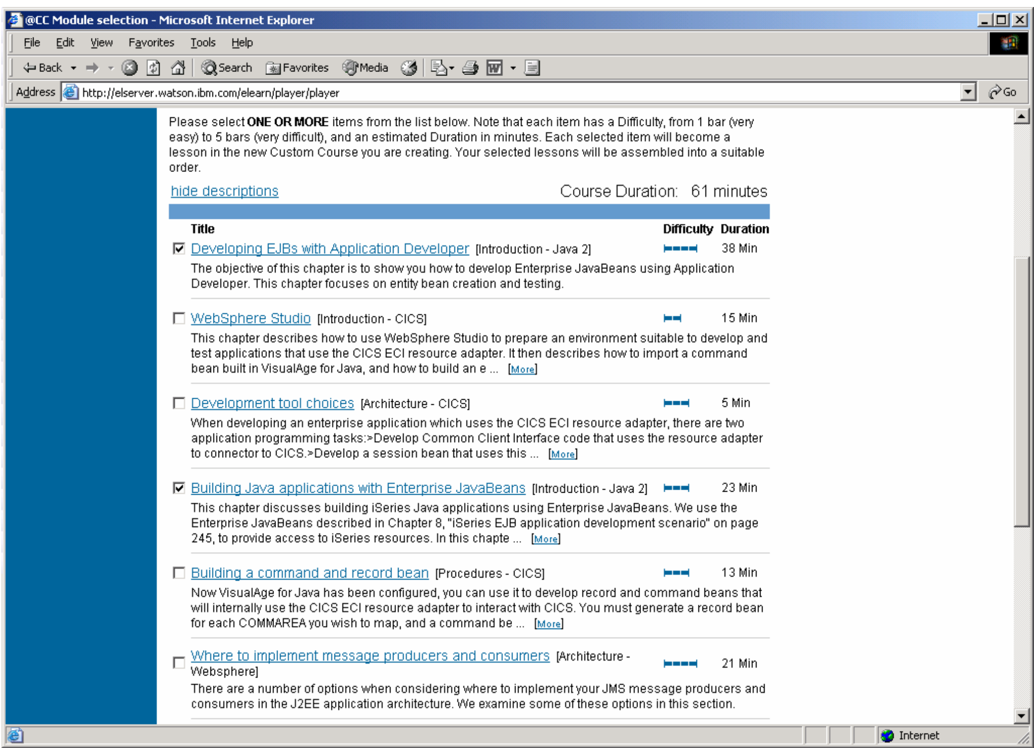

Figure 3: Selecting modules from search results

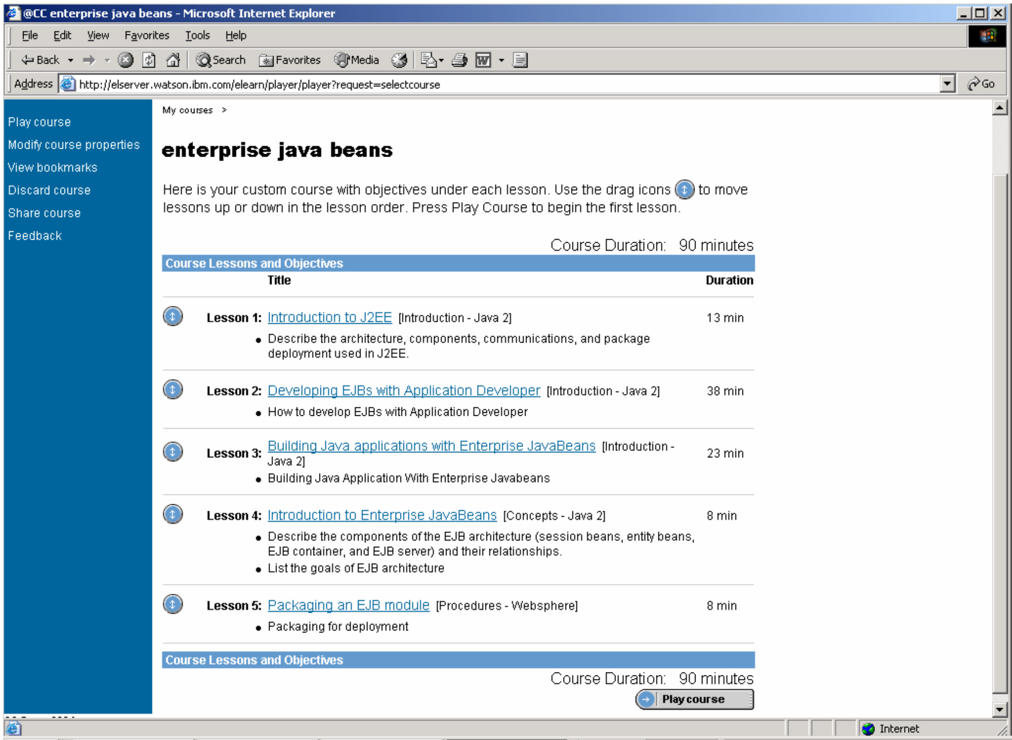

Figure 4: Presentation of a sequence of modules as a custom course

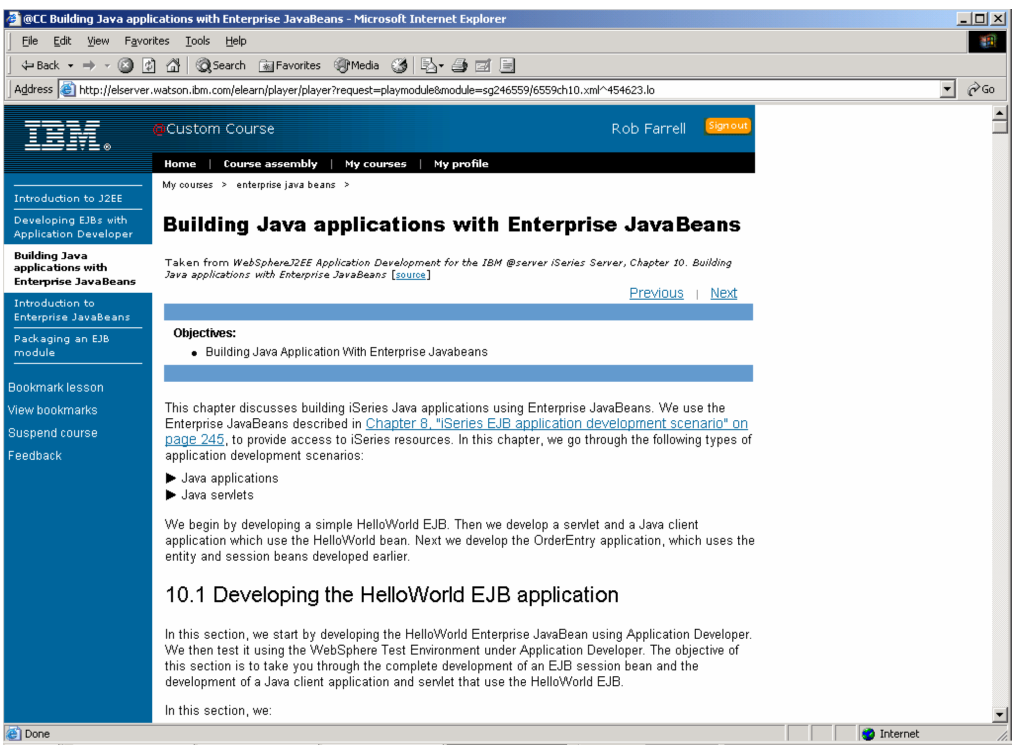

Figure 5: Presentation of learning material as a module within a custom course

Methodology

This section describes the experiment in detail and evaluation methods employed.

Considerations

The experiment was designed with several distinct though inter-related goals. The first goal was to determine whether people could really learn by using the system. We had gone through a number of iterations of user-centered design to assure that the system was fairly *usable* in an objective sense. In addition, we had deployed the system in a fairly extensive field test (Farrell, R. et.al, 2003). Results were encouraging in that users rated the system as highly useable and useful. However, an examination of the server logs revealed that the specific content available to our user groups was relevant only to the specific job-related needs of a small portion of the users and even for those users who used the system extensively for actual work, we had no hard evidence that they learned more effectively with the system than if, say, they had spent the same amount of time using a search engine on the material.

A major challenge in the experimental design was to provide conditions for learning among our set of professional users, all full-time IBM employees. The available volunteers differed considerably in their actual job learning needs and many of those needs were beyond the scope of our available materials. Even if we had had the large amount and variety of material available that would have satisfied all their real learning needs, those needs were so diverse that any meaningful comparison of methods would have been impossible. On the other hand, if we simply required every subject to learn a specified set of materials, while this might provide good experimental control, it would not be very close to the motivational conditions that would apply in the real context for which we were designing our system. Most importantly, a central point of the system is to allow people to select their own materials. Forcing everyone to learn precisely the same material, though tempting from a statistician's perspective, would be beside the point.

Stories are one way to present material in an engaging and motivating manner (Thomas, 1999). Running a series of pilot subjects confirmed that a successful motivational technique applicable here was to introduce subjects to a specific job-related scenario that put them in the position of having to learn something within a general area. The scenario (presented below) was designed to be extremely challenging (to help avoid ceiling effects) but also open to some interpretation of focus so that different subjects would have a great deal of latitude in the materials they chose depending both on what they alread knew and how they perceived the situation. Yet, since they were all reacting to the same scenario, it was hoped that there would be enough commonality of material to allow for some reliable statistical comparisons to be performed.

After learning, subjects were asked to list requirements and to present a high level design. Previous work on design problem solving had revealed that open-ended design problems, while often quite motivating and providing ecological validity, tended to produce extremely diverse results (Thomas & Carroll, 1978; Malhotra, Thomas, Carroll, and Miller, 1980). For this reason, the experiment was designed to capture a variety of different dependent measures.

A second major goal of the experimental design was to provide on-going feedback to the design team so that any remaining bugs, usability glitches, suggestions for improvements or additional functionality could be considered for implementation as early as possible in the overall system development process. Ideally, such a goal, in itself, may be well-served by having users be required to exercise all the functions of the system and observing them face to face or behind a one-way mirror. However, it was deemed even more important to employ users who were representative of our target users; that is, in this case, IBM professionals from a variety of organizations and locations. Therefore, most subjects were run "remotely" (at their normal workplace) and not forced to use all the functionality of the system, but observed to see how they used whatever functionality seemed to them most likely to help them solve the problem presented in the scenario. Arguably, this provides a much more "ecologically valid" look at user behavior than a controlled experiment (Thomas & Kellogg, 1983).

A third more general and exploratory goal was simply to learn more about the general behavior that occurs in self-directed, on-demand e-learning. The richness of the scenario and related design problem helped fulfil this goal. There were also an additional number of subjective and objective measures that were collected and analyzed in the service of this more general goal.

Experimental Design

The experimental design was fairly straighforward. There were two main manipulated conditions. Approximately half the subjects were randomly assigned to the "Query Only Condition" and about half were assigned to the "Custom Course Condition." All subjects were given the same general instructions, the same background questionnaire, and asked the same "objective" factual questions about web services and WebSphere, given the same scenario. All asked to produce relevant requirements and a high level design. Almost all subjects were run remotely by a combination of conference call and having the experimenter and subjects logged into the same web site so that the experimenter could generally follow the subject's behavior remotely. All subjects were given the same subjective questionnaire and asked for free form comments.

In the "Query Only Condition," subjects had access to the first portion of the system; that is, they could use the search engine we provided to enter query terms they thought relevant to learning about how to solve the problem presented in the scenario. They were free to browse through the materials returned, take notes, cut and paste, and enter additional queries.

In the "Custom Course Condition," subjects had the system construct an assemblage of learning objects based on the query results that they selected. Subjects were again free to go through those materials in any order, take notes, cut and paste and to go back and construct additional "courses."

In somewhat more detail, the interface and functionality experienced by subjects in the two conditions, though very similar, differed in several important details. First, what the subjects ultimately looked at in the "Custom Course Condition" was referred to as a "course" (as opposed to "results"). Though this extends the term "course" beyond its usual meaning in the educational literature (and we now refer to it as a "Learning Environment"), the term "course" may have helped induce an expectation of learning. Second, there were stated learning objectives for each module in the "course" which might help serve to focus the task of reading the materials. Third, subjects in the

“Custom Course Condition” were shown additional metadata (description, difficulty, duration for each module) which may have helped them plan more effectively or helped set expectations. Fourth, though subjects in both conditions were free to navigate the material in any order, in the “Custom Course Condition,” the system rearranged material in a pedagogically coherent order. Lastly and perhaps most importantly, subjects in the Query Only Condition were not allowed to select modules and were not offered an “Assemble”button to create a custom course, whereas subjects in the Custom Course Condition were allowed to select modules and press “Assemble” to create a custom course. Thus, the major interaction modes in the two conditions were quite different. The Query Only condition was similar to a typical search engine, while the Custom Course condition afforded additional steps of deciding which search results were relevant based upon the available metadata and asking the system to rearrange the selected, relevant search results into a default sequence that served both as an advanced organizer for the material in the custom course and as a somwhat “personalized” navigational structure to each module.

Procedure

Volunteers were solicited by our IBM Global Services partners and told that the experiment would take about two hours and would involve learning some material via an on-line e-learning system. Subjects were scheduled by mutual agreement for a continuous two hour session. Shortly before the experiment, they were e-mailed the experimental materials and the web address to access the Custom Course system on our server. Two subjects were run by a combination of e-meetings and telephone conference. Two subjects were run face-to-face. All other subjects were run by phone with the experimenter signed in to the system and attempt to “shadow” the subject’s behavior. All subjects were encouraged to “think aloud” during the session and the experimenter attempted to capture the essence of these comments in real time.

Occassionally, the subjects would ask the experimenter for substantive advice about the solution to the problem or where to look for information. These questions were deflected as being outside the area of the experimenter’s expertise. However, any questions about the nature of the experiment or the experimental procedure were answered.

Subjects were free to use whatever off-line or on-line tools they felt most comfortable with to provide their answers and solutions (pencil and paper, presentation software, word processing software, etc.). However, they were asked not to look up answers in books or ask expert colleagues for advice. It was also made clear that results would be anonymized and that the point of the experiment was to learn more about e-learning and not about them or their abilities.

Subjects

In this paper, we typically refer to the participants as “subjects” when focusing on their selection and the experimental design, as “users” when focusing on the interactive aspects of the system, and as “learners” when focusing on the pedagogical implications of the study. The original plan was to run 15 to 20 subjects in each of the two experimental conditions. For various reasons only 26 were run. In a few cases, subjects failed to complete one phase of the experiment and hence, the degrees of freedom differ slightly for different measures.

The majority of the subjects were classified as "IT Specialist"; some were consultants, one IT architect, sales specialists, managers, and some "others." There was a great deal of variability of the breadth and depth in their general programming background as well as their pre-existing knowledge about web services and WebSphere.

Background Questionnaire

Subjects were asked, as part of the two hour experiment, to fill out a brief questionnaire on their background. The questions included years of experience in the IT industry, years of programming experience, number of languages in which they had written a program of more than 100 lines, number of langauges in which they had written a program of more than 1000 lines, self-rated competence in web services (0-5) and WebSphere (0-5). The two groups were roughly comparable on these measures but there was considerable variance within each group. For example, the mean number of years in the IT industry was 19.5, but varied from 2 to 30. The number of years as a programmer varied from 0 to 30 and the self-rating on WebSphere competence varied from 1 (very minimal) to 4 (competent to use for customer).

Subject Tasks

Subjects began by reading an overview of the experiment and were asked if they had any questions. They then read the scenario below to help motivate them to design a solution to a customer's problem. Subjects then had 55 minutes to construct and use one or more "Custom Course" or to use the search engine alone in order to learn what they needed to know to do a high level design. Subjects then had 35 minutes to complete a high level design including a list of requirements. It should be noted that this design task would normally take far more time. Because we thought it would be instructive to see how much *could* be accomplished under foreshortened conditions. Subjects then answered five "objective" questions about Websphere and web services, answered subjective questions about their experience with the software (on a five point Likert scale), and were also given a background questionnaire. Finally, they were encouraged to write any additional comments in free form about the system or their experience.

Scenario

Users were presented with a scenario. At the conclusion of the experiment, they were told to:

"construct a custom course that will allow you to learn as much as possible in a little less than an hour so you'll be able to carry out the design briefing "on your feet" in front of the customer using only brief notes".

The scenario was fictional, but was based upon several real scenarios presented in the books that were part of the learning material users had availble to them. Here is the scenario in full:

IBM has the chance for a moderately large software deal with Genysis although {Real company and product were named here for the users} product is a strong competitor.

You've been called in at the last minute because the person who was supposed to give a client presentation was in a serious automobile accident. The client exec wanted to reschedule the meeting but Genysis says that they need to make a decision now. You will meet with the client in about an hour and be asked at that point to provide a preliminary and high level "design" to solve their problem. If you can pull this off, your boss would be extremely impressed. Basically, their situation is as follows:

Genysis is an international provider of herbs, vitamins, and related health care products; they have a working client-server application that allows registered pharmacists to order their products on-line. A module of their software called "Doser" allows the pharmacist to enter a body weight and get a recommended dosage for particular drugs. A series of recent events, including the results of an extensive marketing survey and some recent changes in legislation have led them to want to publish "Doser" as a Web Service available to the general public. Moreover, because the marketing survey indicates that a large number of their potential customers in foreign countries do not speak English, they would like to offer interactions in at least six major languages: Mandarin, French, German, Italian and Spanish as well as English. Luckily, although the current application is written only for an American English bias, the Model View Controller Pattern was used and formats, help files and messages are all separate from the logic of the control flow. Genysis do not themselves have expertise in these various languages; however, they are considering the possible use of another Web Service called PolyGlot to do the translation for them.

Genysis would like the end customers who visit their website to be able to choose a language, choose drugs and weight ranges from language-specific pull-down lists, press enter and have returned language-specific dosage recommendations.

Currently, Genysis is using DB/2, JAVA application programs on the server and a small but specialized client program. They want to move to a solution in which the only requirement would be for the end user to have a standard browser and internet access.

Subjects were also given written instructions for what they need to produce as output following their study:

At an upcoming customer briefing, you're supposed to list what you see as requirements ---- a list of functions that the solution should provide. Second, you need to create a high-level architectural diagram showing the major parts of the system, how they interact, and which functions reside in which parts. You should include any EJB's, servlets, JSP's, JAVA classes, service beans, and associated calls and show what Genysis needs to build the service, what's involved in accessing the service and show what happens at run-time. In an hour, obviously, you cannot detail precisely everything, but given the time constraint, try to be as detailed as possible.

While the main point of this experiment was to determine the effectiveness of the Custom Course system, it should be noted that the process of using a highly motivating scenario was highly effective in motivating the subjects to perform searches and attempt to learn. This scenario resonated strongly with important aspects of the subjects' real lives. It should be noted that our use of scenarios for motivating a learning task is compatible with, but distinct from the growing use of scenarios as tools for the design of Human Computer Interaction (Carroll, 1995).

Design Evaluation Methods

Grading the various subject tasks was fairly straightforward except when it came to evaluating the designs. Since design is often a subjective process with a wide variety of "right" answers, a number of design evaluation methods were used. First, a number of quantititative measures were taken. These often correlate with the more elusive qualitative measures. That is, subjects who engage in more overall design behavior tend to have better designs as well.

From this perspective, designs were quantified in several ways: the total number of pages of output, the number of words, the number of boxes, and the number of relational arrows (not including items simply cut and pasted) as well as the sum of words, arrows, and boxes. In addition, one expert in the field drew a "solution" to the problem and then each subject's design was graded in terms of the simple presence or absence of each feature.

In addition to these quantitative measures, three experts in both web services and Websphere rated each of the designs in a manner that was blind to condition.

Results

This section describes both the quantiative and qualitative data collected during the experiment and the results we obtained.

Quantitative Design Results

The Custom Course group produced more design behavior than the Query Only group. In terms of total pages of output, the Custom Course group produced an average of 6.5 pages as opposed to 1.83 for the Query Only group: $t(24) = 2.1$, $p<.05$. There was also a significant correlation between self-reported experience and the number of pages; $r=.46$, $t(24)=2.53$, $p<.01$ (one-tailed). If we "correct" for this effect by looking at residuals, the effect of group increases: $t(24)=2.66$, $p <.014$.

Experience also correlated with other quantitative measures of design. The correlation of self-reported experience with the total number of words, boxes, and arrows was $r=.409$, $t(22)=2.1$, $p<.05$. If the scores are "corrected" for experience by taking the residuals in this correlation, the Custom Course group outperformed the Query Group, $t(22)=2.59$, $p<.02$. Experience also correlated separately and marginally with boxes, ($r=.418$, $p<.05$), arrows ($r=.346$, $p<.1$) and words ($r=.367$, $p<.08$). With experience accounted for in this way, the Custom Course Group also had more words $t(22)=2.46$, $p<.03$. The self-ratings on experience also correlated with performance on the so-called "objective tests" of knowledge of web services and Websphere. These latter consisted of five factual questions, a question about the three main standards associated with web services and a question about the logical sequence of developing and deploying a Web Service. Here the correlation was $r=.422$, $t(23)=2.23$, $p<.04$, two-tailed.

As mentioned above, an expert in web services and WebSphere was also asked to examine the scenario and produce a set of requirements and a high-level design. Then, each of the designs was graded according to the number of features present or absent when compared to that presumably correct design. In this case, the mean number of features in the Custom Course group was 10.8 but only 8.6 in the Query Only group. This was not statistically significant. Again, the effect of group was dwarfed by experience. The correlation between self-reported experience and number of features was $r=.463$, $t(22)=2.34$, $p<.05$. If this is taken into account, the difference between the two groups becomes statistically significant $t(22)=2.35$, $p<.03$.

Some example designs from the two conditions are shown below in Figures 6 and 7.

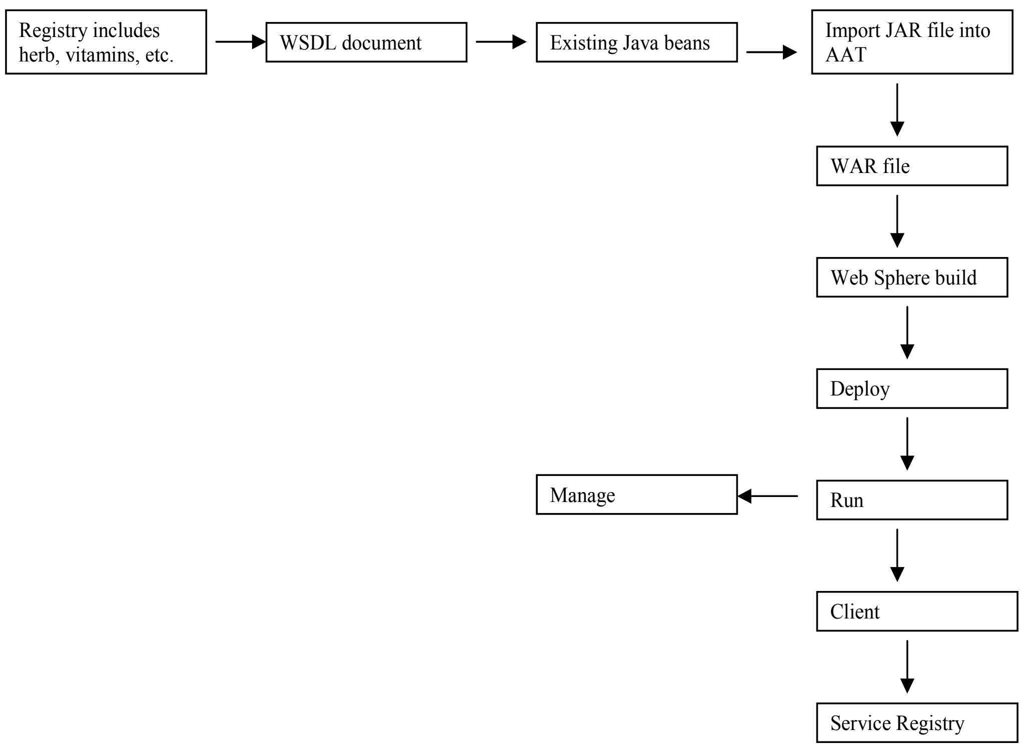

Figure 6: A Sample Design from a subject in the Query Only condition

Requirements

- WebService enabling existing Business Logic
- Multilanguage (Mandarin, French, German, Italian, Spanish and English)
- **Business Logic of existing solutions runs on server**
- **New solution runs exactly the same Business Logic**
- **Content of the database is language independent (product names and numbers)**

New Architecture

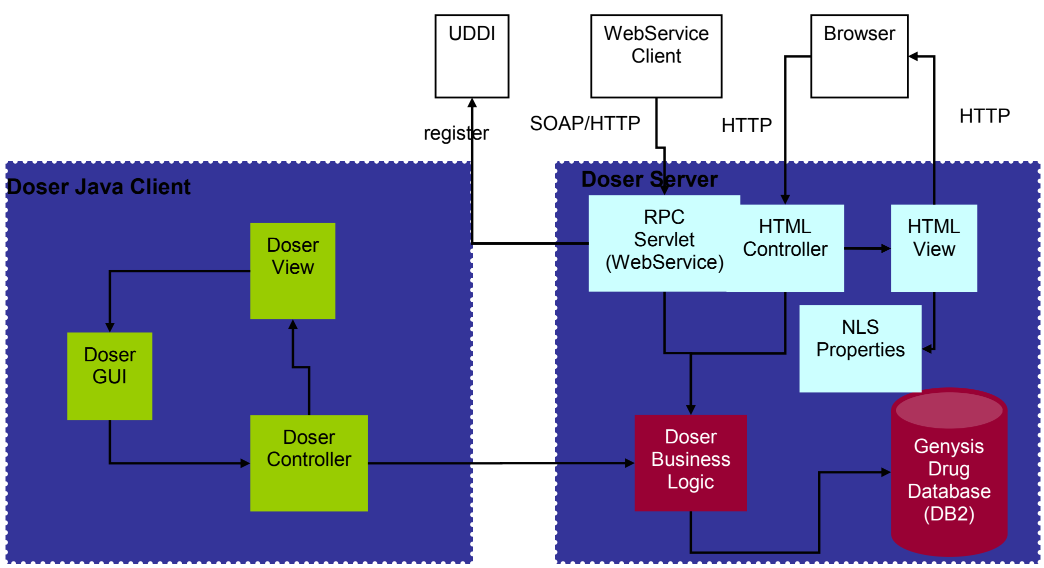

Figure 7: A sample design from a Custom Course condition subject (p 5)

Qualitative Design Results

Designs were also graded independently, and blindly as to condition, by three experts. Each expert had an advanced degree in Computer Science and considerable hands-on experience with both web services technologies and web application design and development. Each design was rated (A, B, C, D, or E) on Presentation, Level of Detail, Accuracy, Completeness, and Depth of Understanding. In addition, an overall rating was assigned as equal to the sum of these five dimensions. Graders tended to agree overall on overall quality: $r(1,2)=.728$, $t(23)=5.092$, $p<.0001$; $r(1,3)=.72$, $t(23)=4.98$, $p<.0001$; $r(2,3)=.723$, $t(23)=5.183$, $p<.0001$.

Looked at separately, there was substantial agreement among the three raters separately on all the subscales as well. For rater pairs (1,2), (2,3) and (1,3), these correlations were respectively: Presentation (.27, .68, .39); Detail (.56, .65, .56); Accuracy (.47, .54, .43); Completeness (.64, .54, .40), and Depth of Understanding (.60, .73, .52). All these were statistically significant at least at the .05 level except for raters 1 and 2 on presentation. Of course, it is not surprising that the correlations tended to be somewhat lower for the subscores than the overall scores since the latter are based on more data and the raters were not trained on distinguishing these subscales. Such generally high and significant agreement is reassuring because design problems, by their very nature, admit of multiple potential solutions. Even in IT, it is unlikely that different designers, given the same situation, would produce exactly the same designs. Nonetheless, given the wide range of quality in designs in this experiment, there was basic agreement on their relative merits. **The subjects in the Custom Course Condition produced "better" overall designs in the opinion of each of the three raters, but not significantly so. In adjusting for experience (total design score – total experience score), more "good" designs remained (6) for the Custom Course group than the the Query Only Group (1)**; Chi Square (1)=4.43, $p<.05$.

If we look at the various subscales separately, we find a similar pattern. The Custom Course Group did better on each of the five subscales, but not significantly on any one. A one-tailed probability that all five tests would turn out in favor of the Custom Course Group would be .03125 (1/32), but this calculation would be based on the dubious notion that the scales were really rated independently and generally raters exhibit a "halo effect." Self-reported experience was also positively, but non-significantly correlated with scores on all five subscales. If one looks at "unusually" high or low residual scores; that is, scores where quality – experience was unusually high (>2.0) vs. unusually low (<-2.0), there were significantly more positive (rather than negative) residuals for the Custom Course Group compared with the Query Only Group (Chi Square (1) = 7.79, $p<.01$). Looked at in this way, each of the five subscales had the same trend; however, only the Accuracy subscale was statistically significant, Chi Square(1)=6, $p<.025$.

Subjects with higher "objective" scores also tended to have better designs: $r=.598$, $t(21)=3.41$, $p<.003$. It appears that subjects from the Custom Course Group overall produced more in terms of design and, with experience accounted for, qualitatively better designs as well. Individual differences in relevant experience were large and tended to be better predictors of design quality. Subjectively, and very roughly, it appeared that there were really three distinct cases. **A few subjects already came into the experiment with most of the knowledge they needed to do the design. For these subjects, which system they used made little difference. They basically already knew enough to do a credible design and needed to find the answers to a few details. Another, larger group of subjects was hopelessly unprepared in terms of background. For them, an hour was far too little time to learn enough to produce a credible design. An interesting intermediate case were subjects who had some relevant background knowledge. And, for these subjects, it appeared that the Custom Course functionality really did allow them to learn enough in about an hour's time to produce better designs than their counterparts in the Query Only condition.**

Custom Course Diversity

The result that the Custom Course system software group seemed able to make generally better designs than the Query Only group was of primary importance, a more detailed look at the behavior of subjects during their interactions with the software was also of interest. The premise of dynamically assembled learning objects is that the somewhat knowledgable adult user will be able to use the software to assemble material that is sensitive both to his or her own background and to the situation at hand. If this goal was met, then the behavior of subjects in the experiment should reflect two things. First, the learning objects that were ultimately chosen should be relevant to the scenario given. Second, because different subjects had different backgrounds, there should be considerable diversity in the learning objects chosen and visited.

While we have developed no standardized quantitative method for measuring the diversity of courses or the diversity of backgrounds, a few examples may illustrate both the diversity of selected learning objects and the sensitivity to background as well as the relevance to the scenario. A sales specialist who was novice with respect to WebSphere and web services asked a single query based on these keywords: "Web services, language translation, Model View Controller Pattern, PolyGlot." This person then assembled a single Custom Course from the query results and studied these learning objects for roughly an hour. These included: "Portal Solution Guidelines", "Web Services introduction", "Model-View Controller Pattern", "publishing and exploring a UDDI registry", and "web services development overview". Notice that the choice of keywords here matches exactly keyword information stated explicitly in the scenario. This is not surprising because the novice has little basis for translating into other terms.

By contrast, an IT specialist with little experience in web services or WebSphere used only one query, "Java" and assembled a single course from the query results and studied four modules listed: "Application developer terminology", "Creating Java applications", "Creating a Web service from a JavaBean", and "Technology options". Again, these modules are relevant to the scenario but are reflective of this learner's higher fluency with the IT area.

A consultant with no background in web services or WebSphere again used only one query: "Web service" and visited three different learning objects: "Web services development overview", "Web services creation using the technology preview," and "Creating a web service from a JavaBean". A project manager with some background and knowledge of web services and WebSphere asked four queries and proceeded to use three of these for building associated courses. The three queries used were: (1)"J2EE DB/2 WebService WebSphere portlet EJB JSP" (2) "How to write a webservice" and (3) "WSDL example". Across the three courses, a fair number of learning objects were visited: "Portlet development guidelines", "Preparing for WebSphere application migration", "Foundation and tools", "WebSphere V4.0 assembly and deployment tools", "web services application design and architecture", "Web services introduction", "Introduction to WSDL", "Creating a web service top-down from WSDL", and "Using Appicaiton Developer Integration Edition". It seems clear that this person's familiarity is reflected both in the choice of advanced material as well as his ability to visit such a large number of separate learning objects in the time provided; in this case a total of nine different learning objects were looked at as opposed to 3,4, and 5 for the learners who had been unfamiliar with web services and WebServices. Still more extreme was the case of an IT specialist with 21 years experience in the IT industry and prior knowledge of WebSphere and web services. He issued four queries: (1) "web services", (2) "language translation", (3) "WSDL" and (4) "architecture diagram" and from these visited a total of twelve learning objects: "Web services introduction", "web services development overview", "web services scenario": "architecture and implementation", "multilingual web services", "machine translation", "locale model", "a development methodology for globalized applications", "introduction to WSDL", "creating a web service top-down from WSDL", "sample business process" and "web services gateway".

The learning objects were taken primarily from IBM Redbooks, technical guides created by a team after a product, service, or technology is first distributed. There was a non-significant correlation of visits with size ($r=.098$, $t(8) = .28$, $p>.7$). The experimental instructions included: "Your queries can cover such topics as WebSphere, web services, and globalization." The Redbooks visited in the experiment included seven with "WebSphere" in the title; one with "web services" and one with "Globalization."

In the experiment, there were twenty-nine Learning Objects which were visited by only *one* subject each. *No* Learning Object was visited by *every* subject. There were some relatively popular Learning Objects visited by 11, 11, 7,7,7,6,5,4,3,3,3,3 times each. **Every one of the Custom Courses constructed was somewhat different but every one of them was clearly relevant to the scenario. It seems clear then, based on this qualitative evidence, that people were capable of chosing material that was both relevant to their current level of knowledge and the situation presented.**

Behavioral Differences during the Experiment

Design outcomes were different between the two experimental groups. To look more closely at evidence of differing cognitive processes, we also examined whether the behavior of the subjects differed *during* the experiment. One measure of behavior revolves around the recorded verbal behavior of subjects during the experiment. In particular, the Custom Course Group used significantly more words relating to cognition; e.g., "learn," "know," "creat*," "forgotten," "studying," and "think." The ratio was 74/793 words in the Custom Course group as opposed to 32/803 in the Query Only Group: Chi Square(1)=17.54, $p<.0001$; Fisher Exact=.000). On the other hand, the Query Group had more recorded words that had to do with navigation. Words such as "search," “get,” and “find” constituted 54/803 as opposed to only 13/793 in the Custom Course Group (Chi Square (1)=28.408, p<.0001; Fisher Exact = .000). **This vocabulary distribution may be indicative of the cognitive experience of the subjects: The Query Only users concentrated on searching; the Custom Course group on learning.**

This navigational focus in the Query Group is also reflected in their higher number of separate queries recorded during shadowing. The mean for the Query Only group was 4.71 and only 2.1 for the Custom Course Group (t(15)=2.16, p<.05). Since the above data are based on what the Experimenter recorded during sessions, it is conceivable that an experimental bias might have contributed to the difference in recorded words. For many subjects, we also have Server Log Data detailing what queries were made and these tell a similar, if somewhat more complete story. The mean number of queries for the Custom Course group was 5.0 and the mean number for the Query Only group was 12.5: t(20)=2.16, p<.05. Again, in terms of Sever Log Data, the mean length of queries in the two groups were similar (2.68 for the Custom Course Group and 2.34 for the Query Only Group). Roughly, the two groups used similar terms with similar frequency distributions. However there was a slight difference between the groups in terms of the frequency of visits to Learning Objects from various sources (Chi Square (9)=21.42, p<.025). The main difference was that the Custom Course Group had proportionally more visits to the WebSphere 5.0 book and the "Enabling web services for i-series” book and fewer to the Globalization book.

6891 WebSphere V 5.0	52	52
6176 WesSphere V 4.0	14	27
6851 Globalization	15	58
6585 WSAD	6	16
0192 Enabling Web Services for i-series	16	20
6559 WebSphere J2EE	7	11
6550 WebSphere implementation/integration	2	10
6869 Portal Pattern WebSphere	7	14
0206 CICS WebSphere	0	1
6521 Migrating WebSphere	4	4

Table I: Visits made to various Redbooks by the Custom Course vs. Query Only

The Server Log Data also reveal that the Query Only Group accessed more Learning Objects. Users in the Custom Course Group visited a mean of 9.83 Learning Objects while those in the Query Only group visited a mean of 21.7 (t(19)=2.82, P<.02). Learners in the Custom Course Group visited a mean of 7.5 *unique* Learning Objects while those in the Query Only Group visited a mean of 14.4 *unique* Learning Objects, (t(19)=2.64, p<.02). Not surprisingly, since the groups were given equal time, the distribution of times spent in minutes per Learning Object also differed significantly (Chi Square (7)=26.2, p<.001) with people in the Custom Group dwelling longer on fewer Learning Objects.

Although both groups focused their queries on material relevant to the scenario, qualitatively the groups appeared to focus on slightly different topics as indicated above in terms of the Redbooks. A more detailed look at the Server Log Data helps clarify the patterns. In terms of queries, the groups were similar in their emphasis on the language translation issue; the proportion of queries for the Custom Course Group was .22 and .19 for the Query Only Group. However, in terms of Learning Objects related to translation that were actually visited, the Custom Course Group only visited 13/123 Learning Objects versus 52/195 for the Query Only Group (Chi Square(1)=118.9, p<.001). Similarly, only 2/13 of the Custom Course Group visited any Learning Objects having to do with translation while 7/9 subjects did so from the Query Only Group (Chi Square (1)=10.6, p<.01). **A possible interpretation is that the instructions listed material, in order, about web services, Websphere, and globalization. Since the Custom Course Group tended to spend a longer time on the Learning Objects they visited, although they probably intended to visit material on globalization (based on queries), in fact, most subjects simply never had time to implement that part of their plan.**

There were other slight but apparently significant differences in material visited between the two groups as revealed in the Server Log Data. The Custom Course Group viewed 14/123 Learning Objects with "Creat*" in the title vs. 4/195 for the Query Only Group (Chi Square (1)=12.3, p<.001). The Custom Course Group viewed only 5/123 Learning Objects with "how to", "develop*", "implement*" or "build" in the title vs. 21/195 in the Query Group (Chi Square (1)=4.5, p<.05). The Custom Course Group also viewed 37/123 Learning Objects with "Foundation", "Introduction" or "Overview" in the title versus 41/195 for the Query Only Group (Chi Square (1)=75.2, p<.001). **These results are consistent with the interpretation that the Custom Course Group was more interested in actually learning material and the Query Only group in finding answers. While consistent with what we generally expected to happen, these word frequencies are all ad hoc interpretive comparisons. These latter results are also consistent with the somewhat higher frequency of less experienced people in the Custom Course Group.**

A closer look at the verbal behavior made by subjects during the experiment is instructive for implicit suggestions for changes or improvements to the functionality and user interface. Here are some sample comments from the Custom Course Condition.

"This is not unlike what sometimes happens in my job. Even though I'm a hardware person, somebody who is the expert in something built on top says: 'Why don't you just go add that part of the presentation to the customer.'"

“This is great if I had something opportunity based. It could be really useful in that; better than Tool ABC which is too much and the web you just get blanketed with stuff.”

“I think the custom course concept is great...the one drawback I can see is that if you don’t know the right buzzwords and so on...it’s hard to pick the right query.”

“I have to scroll all the way back up. It would be good if this were in its own window [the text] so I always have these [motions to the nav controls on the left].”

“...so now I get to study for the scenario.”

“What the **** are WAR files?”

“...looks like we know what we’re talking about...”

“I think I’ve forgotten....”

“What is UNICODE?”

“I think it will tell me more of what I need to know...”

“...just to see if there is anything interesting there.”

“I’m picking introduction, not necessarily because I want to learn it but to help me figure out how to show it to the customer.”

“...should be able to print.”

“I’d like to have a way to ‘sicky-mark’ things.”

Here are a few interesting comments from the Query Only Group.

“I would just find the guy who wrote this memo...What I think I need to do is find an architect. There was that one chapter that said [who the various roles were in a project].”

“Can I print anything out?”

“I’m still in web services category...”

“I’m going into...”

“This section talks about...”

“Now to web services architecture...”

“I do another one....”

“...going to go back and change my search criteria now...”

“...go back to...”

“Go back to my main search screen again.”

“I assume we’re going to do a three tiered architecture.”

All the above examples were taken from transcribtion, as accurate as possible, during the experiments. However, subjects were also asked to write free form comments at the end of the experiment. These comments tended to be longer for the Query Only Group (mean=60.6 words) than for the Custom Course Group (mean=28.9 words) ($t(21)=2.13$, $p<.05$). The type of comments differed as well; there were seven specific suggestions for improvements in the Custom Course Group and only one in the Query Group. This may not be that surprising since more functionality (and more new functionality) was offered to the Custom Course Group. Here are some representative comments from the Custom Course Group. "...too wordy...", "...not enough pictures and code snippets...", "...couldn't cut and paste into Tool XYZ." "...need...pop-ups to explain the alphabet soup." "THANKS!", "I would have liked a roadmap." "...would have been very useful...", "To me, it would have been interesting to have a system with definitions for keywords." "You might consider a better ordering....", "...not sure I understood....", "...outside my area..."

Here are some representative segments from the Query Only Group. "....frustrating....", "...could be very useful...", "...frustration...", "...find someone...", "[I was] clueless." "...much too complex...", "...didn't know keywords or phrases...", "...make it easier...", "...because I was so unfamiliar with the task."

Subjective Measures

At the conclusion of the experiment, subjects in both conditions were also given a fifteen item, five-point Likert Scale questionnaire to fill out about their experiences. Overall, both groups had favorable subjective ratings. These were collapsed into a combined score by taking twice the very positive ratings and adding the positive ratings and then subtracting the negative ratings combined with two times the very negative ratings. The Custom Course mean was 9.62 and the Query Only mean was 10.8 ($t(21)=.408$, $p>.68$. There were, however, large individual differences ranging from a total score of 24 to a zero. A significant correlation was found between experience level and subjective satisfaction ($r=.601$, $t(21)=3.445$, $p<.0024$ (two-tailed)). This probably indicates, in conjunction with some of the subjects' comments, that the task and materials were too difficult for the inexperienced, regardless of condition.

Results on Objective Knowledge Items

After the learning and the design exercise, the subjects were also asked a number of "objective" questions about WebSphere and web services. These questions were not known to the subjects ahead of time although they might well have run into information relevant to answering them during the course of their learning. As it turned out, the groups did not differ significantly on these items ($t(23)=.299$, $p=.976$) though there were large individual differences within each group. There was, not surprisingly, a moderate correlation across the groups between performance on these objective items and experience level ($r=.422$, $t(23)=2.23$, $p<.02$, one-tailed).

Discussion

This section takes a critical look at the limitations of the experiment in generalizing its results beyond the experimental conditions. It also explores possible explanations for the results and next steps.

Limitations and Caveats

The range of experience in each of the two groups was considerable (and more than had been anticipated). Not surprisingly, the impact of the differences of decades of experience was not undone by an hour's experience with our system. Under these circumstances, it was felt reasonable to attempt to correct at some level for these large individual differences, given that the correlations between experience and performance.

Generalizing to a broader range of materials, tasks, and subjects would be desirable. It may well be that we have focused on one of the areas (albeit a large one) that is most amenable to this educational approach. Nonetheless, it seems likely that these results will generalize when applied to other situations that involve adult self-motivated learners who are somewhat computer literate when they are learning about a technical area in the context of trying to solve a particular problem. Whether they apply beyond that becomes more speculative. Any one of the following factors might make use of the system we tested less desirable: 1) learners who are not motivated to learn; who want to "game" the system; 2) learners who know very little about the field, 3) learners who have no familiarity with computers or how a search engine works, 4) learners whose learning goal involves a fundamental rethinking of their belief structure, 5) learners engaged with a field of study that requires a heavy emphasis on acquiring new sensory-motor skills, 6) learners engaged in a field of study that requires a heavy emphasis on interpersonal, social, and/or imaginative skills; e.g., negotiation, debate, or creative writing. None of these limitations is probably fundamentally insoluble, but each would require additional functionality added to the system and/or the context of use of the system. We will take up some of these issues again in the final section on future work.

In general, we expect that the system and method describe in this study will work well with motivated learners who can adequately use specific features of the task or problem they face to formulate queries -- words or phrases precise enough to be useful in filling in gaps in their knowledge or skills. Adult learners often understand what they need to learn because of feedback from peers, peripheral participation in tasks, or differences between scaffolded and non-scaffolded performance.Given the ubiquity of search engines today, these learners may already try to delve deeper into specific topics related to their interests or tasks by first querying large databases or the Internet. These users could benefit from our approach.

Hypothesized Explanations for the Effect

There are a number of reasonable hypotheses that might singly, or in combination, account for why subjects produced more complete and better designs in the Custom Course condition.

HYPOTHESIS 1: *The course outline with individual lesson objectives provided an advanced organizer for the information studied as part of the Custom Course.*

Users in the Custom Course Condition may have been served better by the course outline with individual lesson objectives. This outline, not present in the Query Only condition, may have provided an advanced organizer for the material that users found and decided to retain, organize, and in many cases, study.

There is considerable research in psychology on the effects of advanced organizers of various types. Deeper learning involves not only comprehending and retaining detailed and complex information, but being able to form new mental schemas and being able to integrate new information into existing schemas for better organization and retrieval. To improve this kind of meaningful learning, Ausubel (1960) believed that it is important to have students preview information to be learned. Teachers can do this by providing a brief introduction about how information is going to be structured. An example of this might be opening a lesson with a statement that provides an overview of what will be taught, an outline of information to see the big picture, or providing objective for each lesson that explain what learners should understand or know how to do after learning the information. Perhaps some of the most compelling demonstrations of the powerful potential effects of advanced organizers come from Branford and Johnson (1973). Subjects given abstract descriptions of a process (such as washing clothes) have a great deal of trouble understanding and remembering those descriptions. Given the hint (by title or picture) that the description is of washing clothes, the abstract description becomes easily understandable. In a similar fashion, our Custom Course subjects may have found the learning objects sensible only in the context of the learning objectives provided by our knowledgeable metadata coders.

Studies have been done to investigate the effects of objectives on learning and have generally found that well-worded objectives can have a positive effect (Rothkopf & Kaplan 1972), (Kaplan & Simmons, 1974). Specifically, Rothkopf and Kaplan's studies found an effect for intentional learning when providing objectives prior to the text. Specifically stated objectives produced more intentional learning than did general objectives. In this experiment, subjects in the Custom Course condition may have learned more and therefore had more material on which to base their designs because the learning objectives constituted a kind of advanced organizer, providing cues for organizing the long-term memory of the learning objects assembled together in the custom course.

HYPOTHESIS 2: *Proceeding through learning objects in order within a course may have been a more logical and coherent learning experience.*

The dynamic assembly algorithm used may have put the material chosen by the learner in a coherent order that served to improve the connections between otherwise disconnected learning material. While this order was probably not as good as one chosen by an experienced curriculum designer, nonetheless, the order is more coherent than what is provider by a search engine. Although learners were free to go through the material in an order different from that suggested, most of the time, most learners did go through in the suggested order. The better linear order may have been one source of superiority for the Custom Course Group.

Order effects are rampant in psychological phenomena; indeed, they are so rampant that many experiments must be designed carefully to "counter-balance" for order effects.

Such effects can be found in visual perception, auditory perception, kinesthetic perception, affective responses, attitude change, reaction time, and problem solving as well as learning. In problem solving, order effects can be so strong that subjects arriving at a specific state from different previous states may not even recognize the state as familiar (Thomas, 1974). Furthermore, such effects exists on many time scales as well. Visual masking effects take place on the order of milliseconds while the effect of learning a natural language first takes place on the order of years. In the pedagogical domain, there are multi-year curricular constraints; e.g., learning arithmetic before algebra and algebra before calculus. Therefore, the importance of order, *a priori,* seems high. Actually demonstrating the superiority of one specific order of presentation over another for a single body of knowledge is not trivial and certainly beyond the scope of this experiment. Empirically demonstrating the superiority of a *method* or *algorithm* for ordering materials, *in general,* across a variety of settings, learners, and materials seems to be an almost intractable problem. Yet, in light of the ubiquity of order effects in human experience, it seems reasonable to suppose that a logical order is more effective even in the absense of specific evidence. A good order of materials can prevent what Carroll and Mack (1984) refer to as "tangled problems" and enable learners to focus on a manageable subset of materials.

HYPOTHESIS 3: *The additional metadata (description, difficulty, duration) presented in the query results page may have allowed subjects to find more appropriate learning material.*

Several subjects commented during the experiment that (in effect) they found this surfacing of metadata to be a great help in their choices of which items to include in their custom course. Since the Query Only subjects did not see this metadata, they may not have chosen material quite so appropriate to the joint constraints of the design behavior required by the scenario and their own current state of knowledge.

HYPOTHESIS 4: *The extra work involved in making a course may have encouraged the Custom Course Group to stay more focused and less scattered.*

Subjects in the Query Only Group "skipped around" a lot more than subjects in the Custom Course Group. They may have been looking to very specific answers to specific questions and when their search did not immediately yield the answers to those specific questions, they may have entered another query. By repeating this process numerous times, however, they may have missed much material that was relevant to other (but not immediately current) questions whose answers were also necessary for a complete solution to the problem posed by the scenario. In the Query Only Group, the cost for skipping back and forth, in terms of time and number of actions taken was small. In the the Custom Course system Group however, more steps were needed to construct a "course." Once a "course" was constructed, as a consequence of some thought and work on the part of the learner, it may have induced the learner to try more dilligently to find relevant information within that course before abandoning it and building a new one.

HYPOTHESIS 5: *The expectation of building and using a "course" may have induced a more reflective cognitive set.*

Although the term "course" is probably an exaggeration in comparison to the time, effort, and human intelligence involved courses constructed by proficient course designers, the prominence of the term on our interface and in the instructions may have encouraged learners to take a more reflective view of the material they were presented with. Since they were taking a "course" their expectations were higher with respect to remembering, thinking, and applying aquired knowledge when being later tested. When you are simply using a search engine, you are just looking for answers that can then be forgotten. At least, these seem to be a reasonable baseline cultural expectation. The verbal behavior of the learners in the two conditions, as well as slight differences in learning objects chosen, indirectly supports this hypothesis.

The hypotheses above could be considered variations on a theme; *viz.*,that the the Custom Course system software created an environment better suited to a more "active approach" toward learning which resulted in a greater depth of processing. To a degree at least, the software was designed to support a constructivist (Piaget, 1952) view of learning and it seems to have succeeded. Generally, constructivist theories of learning postulate that learning is a process driven by the learner instead of being a response to events or situations. In the context of this experiment, though learners were run individually, the scenario, in effect, evoked the social context in which the learners operate on a daily basis including reference to the learner's management structure, competition, satisfying the customer, and being able to adequately communicate the result in the form of apresentation. Thus, the system effectively supported both constructivism and social learning.

Suggestions for Improvements in the Custom Course system

A number of suggestions for improvements of the learning software grew out of comments and observations made during the experiment. Most have been implemented. A short recap of these will be given here. Modifications to functionality and interface have been based on a combination of feedback mechanisms including: a heuristic evaluation by HCI experts; running a number of iterative design cycles over the last few years; and user comments, feedback, surveys, interviews, and Server Log Data of two field studies as well as behavior, comments, and suggestions from this particular experiment (Farrell, R. et.al, 2004).

One major finding from this study, corroborated by user comments, was that the system worked well if one knew enough about the area to enter the right search terms. As a result, we are considering ways of allowing users, especially novices, to browse relevant materials as an alternative to search.

Our original intent was to constrain the search results and/or the course material on the basis of the user's job description. Our original attempt proved insufficiently discriminating so we are exploring additional ways of having users specify their job context and goals.

Based partly on user feedback, we've improved the algorithms for selecting and organizing material using a topic graph to help fill in "missing gaps" in the user's selected search results (Farrell, Thomas, and Liburd, 2004). In addition, material is ordered by topic and secondarily according to rhetorical use. For instance, introductory material on a given topic precedes architectural concepts which precede code examples and procedures to follow. Users can also now specify up front the “type” of material they would like and whether they want an “in-depth” course or an “overview” course. Users are also able to use “drag and drop” to rearrange the lessons within a course

There are several ways in which the system could be extended. Here are some directions for expansion provided by the users engaged in the experiment:

1. add more interactivity or a simulation capability
2. add an assessment capability
3. add a “pop-up” facility for explaining acronyms and terms.
4. add additional kinds of media, such as animations or short videos.
5. integrate the system with facilities for social interaction; e.g., finding an expert.

Any one of these directions presents a number of technical and user interface challenges.

Conclusion

This experiment has shown that under certain circumstances, providing a query facility that is integrated with a selection and sequencing capability is an effective way of improving learning, as compared with providing a query facility alone. We suspect that this kind of facility is also more effective than standard self-paced instruction with web delivery, which typically provides no affordances for active learners to construct their own paths through learning material beyond the confines of a particular course. In particular, users with a course assembly capability can direct their own learning to address gaps in their knowledge and skills, to the extent that they are motivated and understand these gaps.

Our system provides a simple but effective interactive approach to organizing information for learning. It puts the responsibility for finding relevant material on the learner, but moves the responsibility for organizing this material to the system. The system relies on the judgments of experts for both the proper order for learning materials and the decisions for where to break larger books or presentations into independent learning objects.

We are now investigating advanced text mining methods to automatically derive some of the required metadata. We are also greatly expanding the type and range of material being offered to the learners for dynamic assembly. We feel that the system can provide a highly effective tool for boosting comprehension when properly embedded in a larger social and organizational context and when offered as a better way of searching for relevant learning materials to reflectively embellish, accompany, or anticipate task performance.

References

Anderson, J. R., Boyle, C. F., Farrell, R., & Reiser, B. J. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), *Modelling Cognition*, John Wiley & Sons.

Ausubel, D.P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 51, 267-272.

Bloom, J. S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher,* **13** (6), 4-16.

Branford, J. D. and Johson,, M. K. (1973). Considerations of some problems of comprehension. In W. G. Chase (Ed.) *Visual information processing*. New York: Academic Press.

Carroll, J. M. and Mack, R. L. (1984). Learning to use a word processor: By doing, by thinking, by knowing. In J. C. Thomas & M. L. Schenider (Eds.), *Human factors in computer systems*. Norwood, N.J.: Ablex.

Carroll, J. M. (1995). *Scenario-based design: Envisioning work and technology in system development*. New York: Wiley.

Farrell, R., Anderson, J., and Reiser, B.J. (1984) "An Interactive Computer-based Tutor for LISP," Proceedings of the Fourth Annual Conference on Artificial Intelligence (AAAI-84).

Farrell, R., Liburd, S. D., and Thomas, J.C. (2004). The dynamic assembly of learning objects. *Thirteen International World Wide Web Conference,*New York City, May 17-22.

Farrell, R., Thomas, J.C., et. al. (2003). Learner-driven Assembly of Web-based Courseware. *World Conference on E-Learning in Corp., Govt., Health., & Higher Ed. 2003*(1), 455-462.

Farrell, R., Thomas, J.C., et. al. (2004). Personalized Just-in-time Dynamic Assembly of Learning Objects. *World Conference on E-Learning in Corp., Govt., Health., & Higher Ed. 2004*(1).

Gagne and Berliner (1988). The Cognitive Processing of Information. *Educational Psychology*. 4th ed.

Kaplan, R. & Simmons, F. (1974). Effects of Instructional Objectives used as orienting stimuli or as summary/review upon prose learning. Journal of Educational Psychology, 66, 4, 614- 622.

Malhotra, A, Thomas, J. C., Carroll, J. M., and Miller, L.A. (1980) Cognitive processes in design. *International Journal of Man-Machine Studies,* **12**(1), 91-201.

Piaget, J.: 1954, *The Construction of Reality in the Chil*d. New York: Basic Books.

Rothkopf, E.Z. and Kaplan, R. (1972). Exploration of the effect of density and specificity of instructional objectives on learning from text. Journal of Educational Psychology, 63, 4, 295-302.

Thomas, J. C. (1974). An analysis of behavior in the hobbits-orcs problem. *Cognitive Psychology,* **6,** 257-269.

Thomas, J. C., and Carroll, J. M. (1978). The psychological study of design. *Design Studies* **1,*1,* 5-11.**

Thomas, J. C. and Kellogg, W. A., (1989) Minimizing ecological gaps in interface design. *IEEE Software,*January, 78-86.

Thomas, J. C. (1999) Narrative technology and the new millennium. *Knowledge ManagementJournal* **2**(9), 14-17.