Formative Evaluation of Project GeoSim:

MigModel and Humpop Tutorial

by

Kevin L. Stringer

Project submitted to the Faculty of the

Virginia Polytechnic Institute & State University

in partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE

IN

COMPUTER SCIENCE

APPROVED:

H. Rex Hartson, chair

Clifford A. Shaffer

L. W. Carstensen

March 1995

Blacksburg, Virginia

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to express my sincere thanks to Dr. Hartson, my advisor and committee chairman for his continuous support and assistance. Without his patience and cooperation throughout this research it would not have been possible to finish this project. I would also like to thank Dr. Lawrence W. Carstensen for serving on my committee and for his help .

Special thanks go to Dr. Clifford A. Shaffer for his support, comments, and cooperation. His sincere efforts and comments during the implementation of this project were a major help.

I would like to thank George Chin for his helpful discussions and comments. Finally, special thanks to my wife Jennifer for her unlimited support during this period.

1.0 INTRODUCTION

1.1 What is Project GeoSim?

Project Geosim is comprised of simulation software packages. The simulation software covers areas relating to geographic migration and geographic population patterns. Simulation models are built using a variety of control parameters. The information is presented and controlled through detailed graphical interfaces that the users can manipulate. Each software simulation package has a unique graphical interface that is tailored to the purpose of that program.

1.2 Purpose of this Work

The purpose of this project is to enhance the user interface of Project GeoSim. To accomplish this purpose, the evaluator performed formative evaluation on the Humpop tutorial and the MigModel simulation program. The evaluator correlated the collected data into quantitative, numeric data and results, and qualitative, non-numeric data and results. Using Ideal, the evaluator collected quantitative and objective data. Objective data refers to directly observing the user's performance. Ideal aided the evaluator in collecting usability data for formative evaluation.

Another goal for this project is to establish a baseline of usability measures. In order to thoroughly test each interface involved in the study, human subjects performed several benchmark tasks. The human subjects gave informative feedback that help determine problem areas that need further addressing in the user interface

2.0 USER ANALYSIS

For the purpose of this study, Project GeoSim users were selected from one category: Geography students that have taken the introductory courses, but who had not been exposed to any Project GeoSim simulation software. None of the subjects that participated in the experiment had ever used or seen someone use any Project GeoSim software, such as IntlPop.

The users were chosen from the introductory geography course that was taught during the summer session of 1994. The participants were chosen on a volunteer basis and were paid $5 for their cooperation in the experiment which lasted about an hour.

Users were not expected to be familiar with computers or to have used them in school or any work related environment in order for them to complete the benchmark tasks. A majority of the users selected had operated or been exposed to a computer program that employs a mouse. As far as their level of expertise, most users considered themselves to be a novice and only a select few identified themselves as being an average computer user. This set of users is representative of the typical introduction to geography class based on the demographic information that was collected from previous classes.
The demographic form collected their experience with computers and mouse operations, see Appendix D. The students ranged in age from 18 to 22 years of age. The number of years of college completed by each student varied from one to three years. The subject pool was comprised of five female and five male participants.

3.0 EVALUATION

Formative evaluation is a process of enhancing and determining areas that need further development in a user-interface design. This evaluation focused on subjective, quantitative, and qualitative techniques. A subjective evaluation focuses on the users' opinions regarding the interface operations and layout. A quantitative evaluation centers around collecting numeric results and data, see (Section 5.3). A qualitative evaluation gathers nonnumeric data and results, see Appendix C.

There are several steps involved in performing formative evaluation [2]. The methods in which these steps were implemented in this experiment are listed below each step:

- Developing the experiment
  - Derive testing goals (Section 3.1)
  - Plan pilot testing (Section 3.2)
  - Develop subject instruction for Humpop (Appendix E)
  - Develop subject instruction for MigModel (Appendix F)
  - Using an informed consent form (Appendix A)
  - Using a nondisclosure agreement form (Appendix B)
- Directing the evaluation sessions
  - Directing the test session (Section 3.3)
- Collecting the data
  - Using Ideal
  - Using Hi-8 video cameras
  - Performing verbal protocol (Section 5.2)
  - Collecting demographic and user experience (Appendix D)
  - Collecting user interface ratings (Section 5.3)

- Analyzing the data

  Discussing usability problems(Section 4.0)

  Discussing usability specification problems (Section 5.3)

  Analyzing verbal protocol (Section 5.4)

  Analyzing user interface ratings (Section 5.5)
- Drawing conclusions to form a resolution for each design problem

  Making recommendations (Section 6.0)

3.1 Testing Goals

The goal of testing the user interfaces was to find usability problems in students' interpretations of the user interfaces on an initial performance. There were areas in the user interface that needed clarification. It was hoped that the students participating in the experiment would be able to perform all the benchmark tasks without misinterpreting the user interface. Testing would help clarify areas in the interface that needed further refinement. Testing focused on the main features of the user interface. Testing did not consist of exhaustive exercises in which each and every facet of the interface would be addressed.

Pre-test and post-test questionnaires were administered to gauge the experience level of the subjects and to rate the overall user interface. The pre-test gathered demographic information about each user, focusing on their computer exposure and experience. The post-questionnaire gauged the overall user satisfaction on how well the user interface performed to their expectations. For example one of the testing goals was based on the accuracy of the users responses to the benchmark tasks.

3.2 Pilot Testing

Pilot testing was conducted for each interface to make sure that the questions were clear and concise, eliminating any ambiguity. This was an important step, because any ambiguity in the benchmark tasks would have diminished the quality of results taken during the testing session.

The pilot testing was aimed at making sure that all the main components of the user interfaces were addressed. The benchmark tasks that were administered during the pilot testing came from previous computer assignments from the introductory geography classes that used the Humpop tutorial software and the MigModel simulation program as homework assignments. The original computer exercises were altered to meet a one hour time requirement for testing each user interface.

The pilot testing provided a means for gathering time completions for all benchmark tasks involved in the testing. Each testing session consisted of several benchmark tasks. In order to determine what should be considered a reasonable amount of time to finish each task, an average of time for each task performed by four subjects ,two subjects for each interface, established the current level or time precedent.

The pilot testing showed that there were several different approaches that a user could take to find the correct result. The time taken for each tasked varied considerably from user to user. Users more familiar with computing environments adapted faster to the interface and had faster completion times than those with less computer experience.

3.3 Test Session

The test sessions were conducted in an isolated evaluation room. The evaluator and the subject were present in the room at the same time. The evaluator provided no assistance during testing. The sessions were videotaped for review at a later date. Two video recorders captured the events and relayed them to outside monitors for meta analysis. The Ideal research team conducted the meta analysis.

Before any student began the experiment, he/she was asked to sign a nondisclosure form and informed consent form. After the forms were signed by the subject, the benchmark tasks were administered by the evaluator. The evaluator provided no assistance with the subject throughout the test session.

Qualitative, non numeric data and results, and quantitative, numeric data and results, were collected during each test session. Before and after the benchmark tasks, the subjects were asked to complete a written questionnaire. Then each subject was asked to participate in a post verbal protocol session in which the evaluator and the participant talk aloud discussing what problems they were having, what they expected to happen that did not, and what they wished had happen. Critical incidents were captured using Ideal.

Ideal was used to support formative evaluation. Ideal is a usability tool that aids in collecting usability data. Ideal has several usability tools that format and collect data. One of the tools that was used created usability specification tables. A format was available that allowed the evaluator to select options that helped set up the table to the specifications of the evaluator (Section 5.3). Another tool that was used performed data collection. This tool captured the times and errors associated with each benchmark task. Several timers could be used simultaneously and areas for note taking were available for each benchmark task.

4.0 USABILITY PROBLEMS

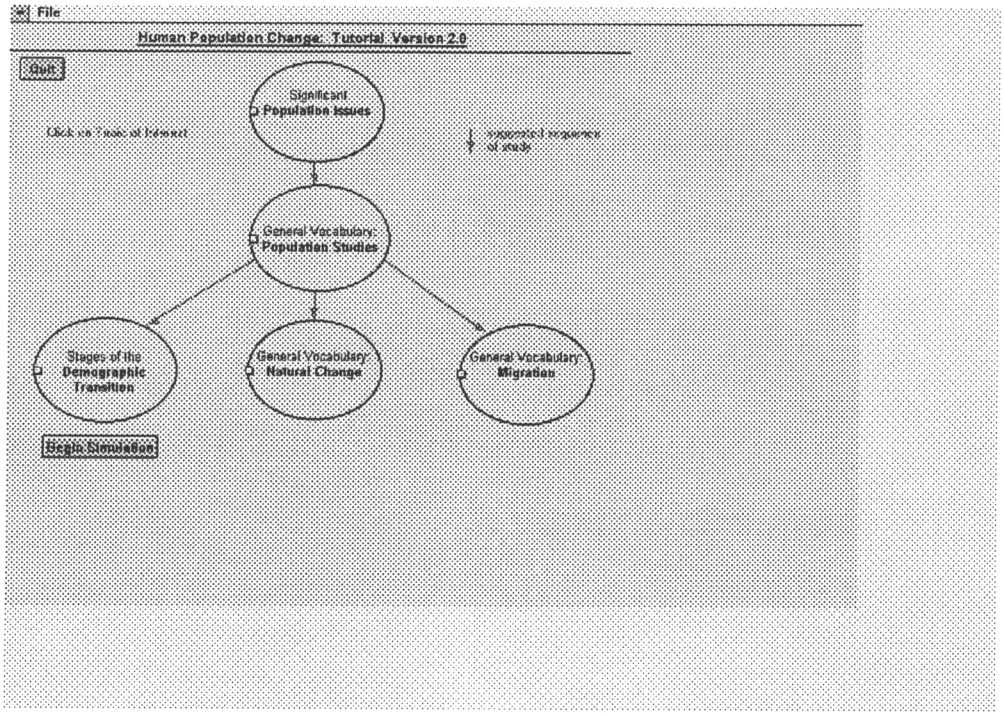

Figure 1: Humpop Tutorial Interface

4.1 Humpop

The Humpop tutorial software is comprised of five sections that discuss human population issues. These issues are discussed using graphics, pictures, and text. A primary goal of Humpop is to familiarize users with human population vocabulary.

The main problem with the Humpop tutorial software was navigation. Users seemed lost throughout the software in trying to perform the benchmark tasks. Users also wanted the capability to go back to previous screens without having to re-enter each unit. The next section will highlight the difficulties with Humpop, beginning with navigation.

4.1.1 Navigation

Navigation is a key ingredient in learning tutorial software. Based on the users' responses, Humpop does not have a good navigation system. A suggested navigation sequence is gven which does not aid the user in learning the system. There was an overwhelming request for users to be able to trace their previous steps. At the present, Humpop does not have any facility that will allow the users to backtrack their steps. The user must exit out of the unit and re-enter at the beginning and traverse through until the desired location is found. A prime example of this is in the Population Issues unit. In this section of the interfaceset of pictures is displayed at a timed pace. If at any point in time the user wishes to stop the display and backtrack to previous pictures or return to the root menu, the user is forced to sit through the entire picture display and then re-enter the unit starting at the beginning.

Many users felt lost at the beginning of the tutorial. Given a set of instructions and tasks to perform many users did not know where to start. The only hint at where a user should start is found in the suggested study arrows that indicate a particular path to follow.

4.1.2 Human Memory Limitation

A good user interface minimizes the working memory limitation on the average user by emphasizing recognition over recall. The user should not have to rely too heavily on their memory abilities in order to gain the maximum functionality from a user interface. The Humpop interface violates this guideline by forcing the user to remember the options that were available in the previous and unsearched Humpop units. Users had difficulty remembering what vocabulary words were contained in each unit. It is only after visiting the units several times does the user recall what contents are in each unit and this is not a successful attempt every time.

4.1.3 System Response Time

System response time can truly enhance the features of an interface or it can cause extreme impatience on the user's behalf. A lot of the system's response time can be directly related to the chosen hardware. In the Population Issues unit of Humpop a display of pictures is shown with related data. The system response time for loading and displaying the pictures was slow for this unit creating impatience on behalf of the users. Users found it annoying to have to sit through each section of pictures.

A solution to the response time could be found by giving the users more options. Most users expressed the desire to be able to skip through some of the slides and sections in this unit. Users that wanted to go back and review previous pictures and information were forced to sit through the entire display of pictures only to have to start at the beginning and wait until the desired set of pictures or information was available. The other solution would be to use faster hardware. The tests were run on a 486 33 which is representative of what the labs were using in the geography department for their introductory classes.

4.1.5 User-centered Design

The Humpop tutorial is not designed around the needs of the user. A good user interface is one that is designed around the users' needs and expectations. One quality feature of a user-centered design is giving the user the feeling that they are in total control of what is about to happen and what will happen. For example, the users are not able to exit out of a set of pictures in "Population Issues" unit. This is a big problem in user-centered design. The system is in control in this case and not the user. Another example in dealing with expectation focuses on the term "continue". It should be replaced with the term "return" in most of the units. The user should not experience the feeling of being lost in the interface. In the Humpop tutorial, users felt lost in the beginning. Users did not know quite where to start looking for the answers to the benchmark tasks. The suggested sequence of study, provided by the arrows, only added confusion and did not disambiguate the situation. Once the user entered into a unit, there was no indication as to where they were in the system. Some indication as to what unit they were in would have helped the situation tremendously.

4.1.6 Consistency

A good user-interface is consistent in screen inertia and is kept as simple as possible. The Humpop tutorial does not have good screen inertia. The commands and menu options were not in the same place. The continue button and return button are placed in different locations throughout the units. Some units allow the user to return to the root menu while other units have no return options. Humpop headings are inconsistent. Some modules indicate which unit you are in by the headings. For example the "Natural Change" unit heading is "Natural Change Vocabulary". For Population Studies, it just has "General Vocabulary" for a heading.

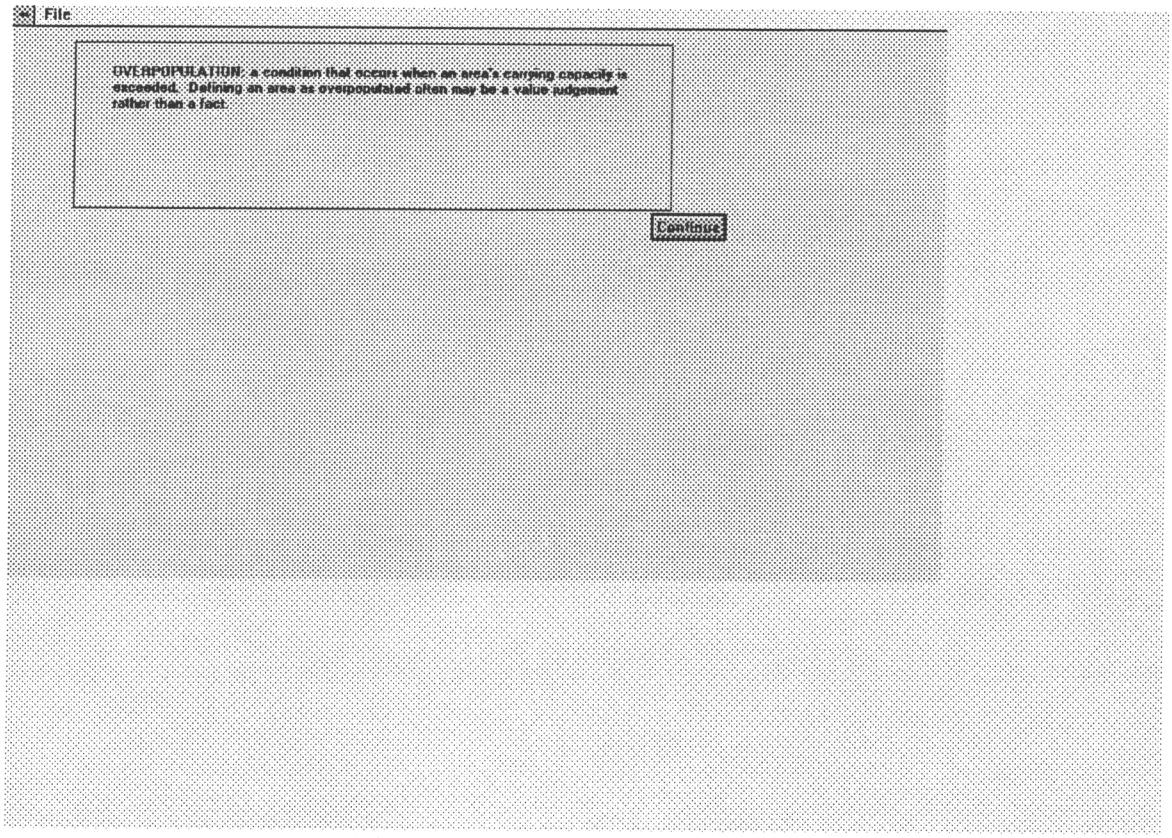

Figure 2: Population Studies unit

4.1.4 Feedback

A system that has good feedback allows the users to gauge the effects of their actions and gives the user appropriate status indicators [2]. One of the shortcomings of Humpop is not knowing in which unit the users are currently working. An example of a violation of this guideline can be seen in any unit. After the users have entered into a unit and select a vocabulary topic the users are left without any visual cues as to what unit they are working in, and the only way to find out is to return to the previous menu or go back out to the root menu. The only indication of where the users were is a colored circle along side each unit. This is not an accurate indication of the unit that the users have come from, because once the users have selected that unit the circle is lit. So re-entering a unit after it has been visited does not change the indication of where the users have just come from. A solution to this problem could be to maintain a history list of the units that the users had already visited.

4.1.7 Simplicity

The operations for a user-interface should be clear in nature. Humpop does a good job in focusing on this guideline. All responses from the user are entered using a mouse. No keyboard entries of any kind are entered from the users. The entire layout of the tutorial is navigated by mouse clicks. The only area that violates simplicity is the number of options that are available on some particular screen that list multiple vocabulary words.

4.1.8 Blocking Errors

A good user-interface prevents users from making errors whenever possible and makes erroneous choices unavailable. Humpop does an excellent job at adhering to this guideline. The users are prevented from entering any erroneous data by only allowing mouse input. There are no choices that require any keyboard input of any kind. This eliminates any erroneous data from being used as input and prevents the users from seeing any system error messages.

4.1.9 System Model

In a user-interface the user should know where they are and where they can go in the system. Humpop tutorial doesn't let the user know where they are in the system once they enter into a unit. For example, once a user has selected a vocabulary word from a vocabulary module, there is no indication as to which vocabulary unit they are currently working in.

4.1.10 Getting User's Attention

One of the key aspects of any interface is to get the user's attention judiciously. This can be accomplished by the size of characters, fonts, intensity, blinking, audio, or color. Humpop does a skillful job at getting the user's attention. The color selection is good. Most of the highlighting is done in blue on a light white background. The unit selections

are in bold black text. The screen layout is spacious in nature and not cluttered. These attributes complement the interface.

4.1.11 Help

Every user interface should be equipped with some form of descriptive help for the user. The main purpose of help is to assist the users when confusion and doubt arise. Humpop presently does not have a descriptive help option that users can select. It would be beneficial if help balloons were inserted in the interface to give more description to the command buttons. This would especially be helpful if they told the users where they would return to and the outcome of selecting certain options.

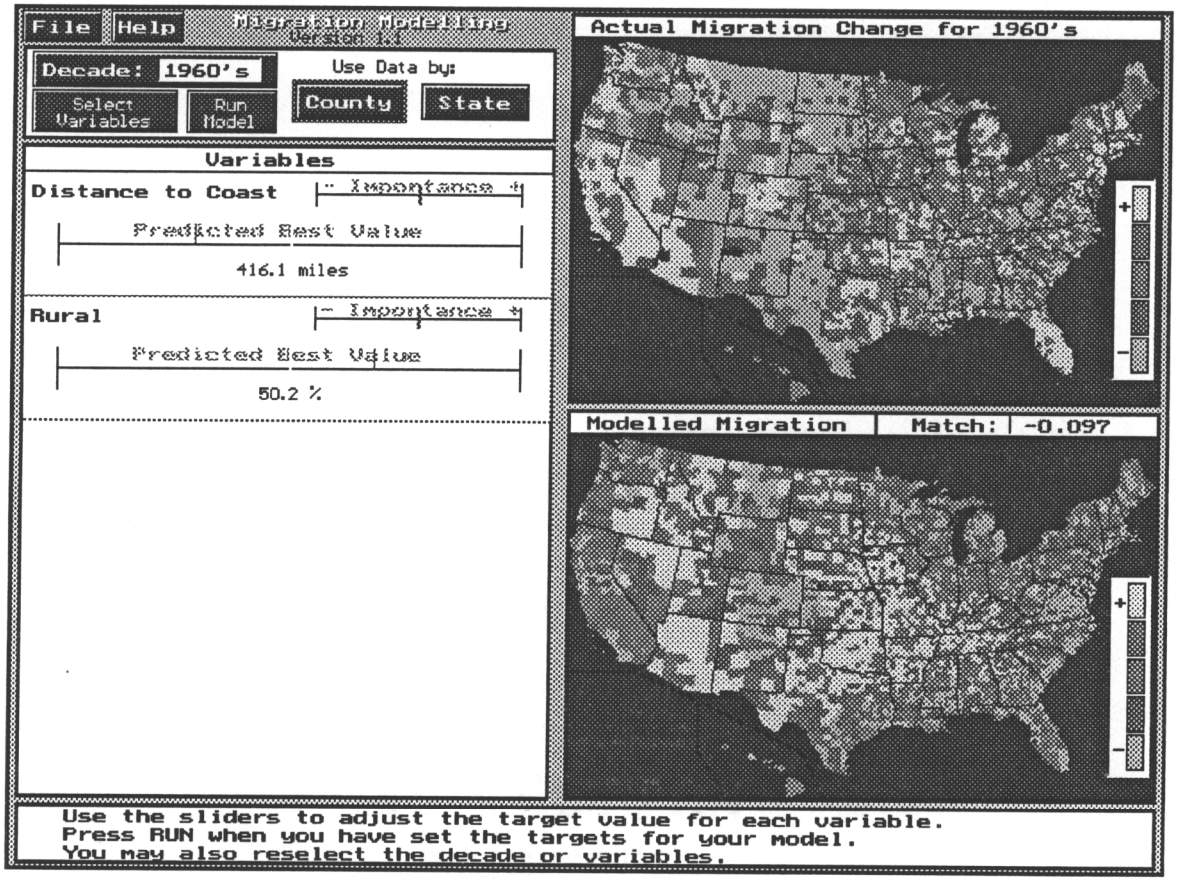

Figure 3: MigModel Interface

4.2 MigModel

MigModel is a software simulation program which allows users to manipulate variables perceived to affect migration patterns. The users are able to select particular decades and adjust the migrations variables. The users see two maps of the United States. The map in the upper right indicates the actual migration patterns for a particular decade. The map in the lower right is the simulated result of variables chosen by the users. Users can manipulate the variable values until they have constructed a map which shows characteristics similar to the original map.

MigModel simulation software has two main problems. One problem centers around getting the user started. For example, users seemed lost at the beginning, not really sure of what their first steps should be. This problem could be addressed by including the tutorial information. The second problem is terminology and graphic representations contained in the user interface. For example, users never really understood what Predicted Best Value meant. The two graphs took upon several different interpretations.

4.2.1 User-Centered Design

A major part of any user interface is to convince the user that they are always in control. MigModel simulation software doesn't give the users a sense of control in the interface. Most users felt lost in the very beginning. The users were never sure if they were doing the right thing or interpreting the terminology correctly. For example, the term "Predicted Best Value" marker was not clear.

4.2.2 Consistency and Simplicity

A user-interface should be consistent in screen inertia and should be kept as simple as possible. MigModel does a very good job in maintaining screen inertia. The screen layout for MigModel only has two areas that change; the Modeled Migration area and the Variables area. The commands are simple and straight forward. The user input comes from mouse selection and not keyboard entry.

The sliders used in the interface are not consistent. MigModel shows one slider when only one variable is selected; a white slider. The problem is that only the white slider moves and the red (mean) indicator remains stationary. When more than one variable is selected, two sliders appear on the screen, each in a different color. A suggestion would be make both movable sliders the same color and indicate that the red indicator is the mean value.

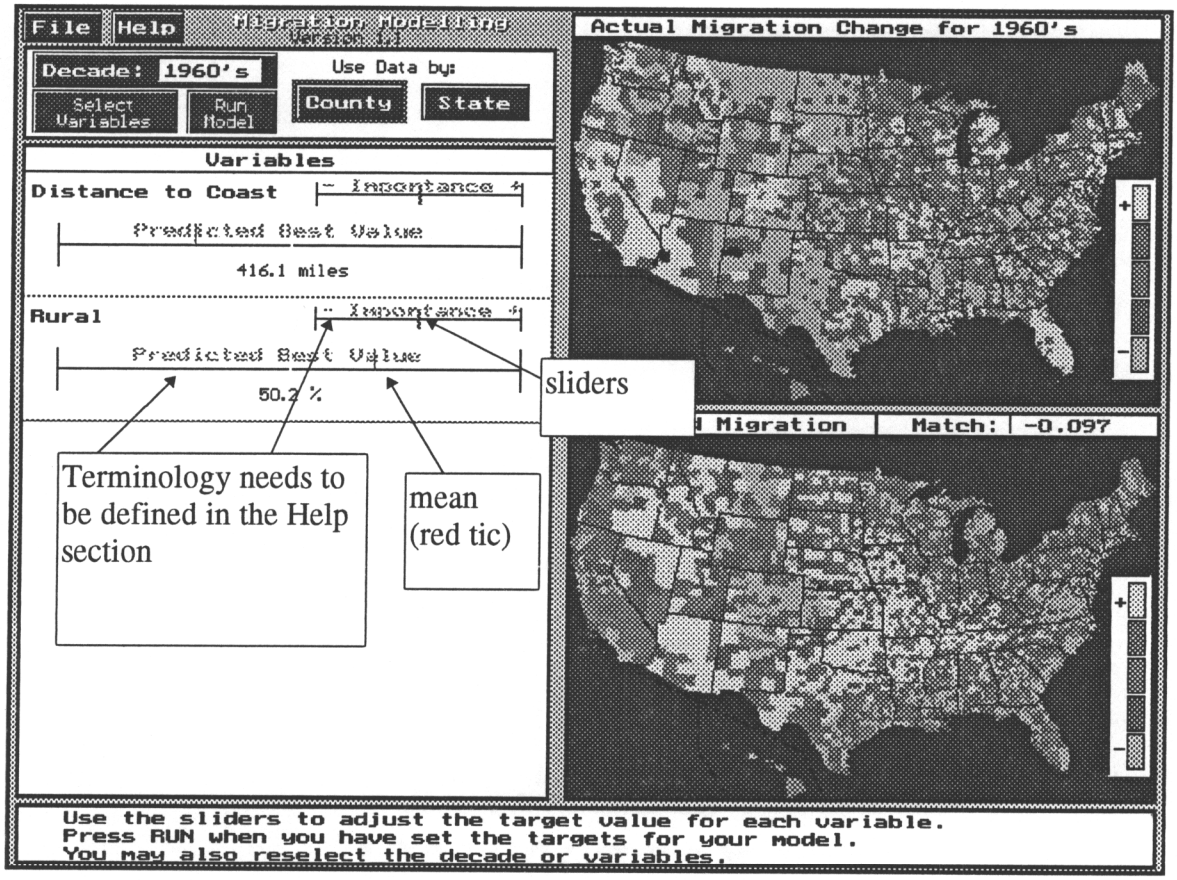

Figure 4: MigModel Terminology

4.2.3 Feedback

MigModel has great feedback when it comes to showing the users how the system has just responded to their requests. Every time a variable is selected it is highlighted in red in the variable window and displayed in the interface immediately after it is selected. When a model is generated by selecting "Run Model" from the interface, a window appears indicating the percentage of model completion. Next, another window is generated indicating the percentage of calculations being done for the match value. Once this is complete, a new lower map is generated from the variables selected.

MigModel has excellent status indicators. The current decade that is selected is visible in the upper portion of the interface at all times. The current level of analysis is also present by a highlighted box around the selected level of analysis.

4.2.4 System Response Time

A key ingredient to user-satisfaction for any user-interface is the response time of the system. The MigModel simulation software package has a large amount of calculations and data that must be analyzed in order to run and draw the models. Some processes take longer than others. For example, MigModel has two levels of analysis, county and state. The county level of analysis is considerably slower and tends to annoy users.

4.2.5 Display Issues

MigModel does a commendable job of organizing the information on the screen. MigModel has a simple layout that is not overbearing or too simplified. The operations were easy because of the highlighting on the screen.

Screen inertia is maintained throughout the interface. Only two areas change throughout the entire use of the interface. The Variables area changes when new or additional variables are selected and the Modeled Migration area changes as new models are being generated.

4.2.6 Cognitive Issues

A good user-interface minimizes the use of cognitive directives (minimizing mental transformation). MigModel violates this guideline by forcing the user to have a general background of the terminology used in the interface. The terms "Predicted Best Value," "Mean," "Match." and the variables in the Variable selection are not defined within the interface. Further, there is no reference to what these terms mean in the help section. Users rely on supplemental material to gain an understanding of what these terms mean.

Many users didn't understand what was meant by high match value. This was left as interpretative material for the user.

Some users confused Predicted Best Value with mean value. It was tempting to align the Predicted Best Value marker with the mean value indicator and expect that to be the Predicted Best Value. Even though it was clearly indicated in the directions that the "red" tic in the interface was the mean value, it was always mistaken as the Predicted Best Value marker.

The graphical representation of both maps was not clear to the users. Again the directions indicated that the upper map in the interface represented the migration pattern for a particular decade and that the lower map was a simulated model. The connection between the two was not clear to the users. Users did not seem to understand that the variables that they were to choose would create the lower map and the goal was to determine what distance or percentage would give the closest match to the upper map. Some were confused because the lower map did not look exactly like the upper map. The users didn't realize that it took several variable settings to get a closer pictorial match.

4.2.7 Blocking Errors

An efficient user interface prevents users' errors whenever possible and makes erroneous choices unavailable. MigModel does an excellent job at adhering to this guideline. The users are prevented from making any erroneous mistakes by only allowing mouse input. There are no choices that require any keyboard input of any kind. This eliminates any erroneous data from being used as input and prevents the users from seeing any system error messages.

4.2.8 Getting User's Attention

One of the key aspects of any interface is to get the user's attention judiciously. This can be accomplished by the size of characters, fonts, intensity, blinking, audio, or color. MigModel does a very good job overall at getting the users attention, but it could use some fine tuning. For example, the terms "Predicted Best Value" and "Importance" appear to be grayed-out giving the impression that they are not options which can be selected.

The color selection adheres to the color guidelines for user-interfaces for the most part. Color guidelines suggest that colors are used to soothe or strike the eye, add accents to uninteresting displays, and to emphasize the logical organization of information [3]. In selecting colors, color pairing is very important. Certain combinations will appear to be garish and difficult to read producing a strain on muscles surrounding the human eye [3]. The pink background seems to produce a strain on human eye and should be replaced with another color, perhaps gray. The other color selections are good. For example, the blue background gives dimension to the maps. Most of the highlighting is done in red on a light gray background which also adhere to the color guidelines. The variable selections are in bold black text which attracts the users attention.

4.2.9 Help

The help options contained in the user interface should be of assistance in clearing up any ambiguities dealing with commands, vocabulary, and procedures. MigModel does not adequately explain terminology and operations that relate directly to the interface. The textual directions concerning model building should be replaced with a graphical representation and more definitions should be added as a separate option in the help section.

5.0 EMPIRICAL USABILITY EVALUATION

5.1 USABILITY SPECIFICATION TABLES

Usability attributes, as target levels for usability, are operationally defined criteria for success of the user interface and not based on the results of the students involved in the study. They are key in determining usability of an interactive system, by providing a metric against which usability of the present version of a user interface can be measured. There are several attributes that comprise a usability specification table [2]. The following were used to gauge usability performance.

Usability Attribute: This is the general attribute that is measured in an interface. There are several usability attributes that can be used to measure an interface. This study looked primarily at initial performance. Initial performance is the user's performance for the very first time using the system. The subjects involved in the study had never seen or used the Humpop tutorial software and the MigModel simulation program.

Measuring Instrument: This is the method used for gathering usability data for a particular usability attribute. This study used benchmark tasks. Benchmark tasks are sets of instructions that a user performs that can be measurable.

Value to be Measured: This is a metric that is used to gather data from the participants involved in the experiment. This metric gathers information from the benchmark tasks.

Current Level: The current level is the value that came from the pilot subjects using the current system version. The pilot subjects involved in the study provided time compilations and responses for the benchmark tasks. These results can be used to show improvement in the user interface.

Worst Judged Acceptable Level: This is the lowest acceptable level of user performance that is considered acceptable for success, not the worst possible answer.

Because we wanted future versions to have improved usability, we set worst acceptable levels at 15% better than the current levels. This figure is admittedly arbitrary, but is consistent with typical situations where current levels are set by an existing system version [2].

Planned Target Level: This is the expected result from the user. The attainment of this target value would indicate usability success. Planned target levels were set to be between the worst judged acceptable level and the best possible level.

Best Possible Level: This is the best attainable level of performance. The best possible level results came from the evaluator's results. The evaluator performed all the benchmark tasks associated with each user interface in the study.

Observed Results: These are the actual values obtained from observing the users perform the benchmark tasks. A mean value of the subjects results were taken and placed in the table.

5.1.1 Usability Specification for Humpop

Table 1: Humpop Usability

Trial	Usability Attribute	Measuring Instrument		Value to be Measured	Current Level	Worst Judged Acceptable Level	Planned Target Level	Best Target Level	Observed in Study
1	Initial Performance	Benchmark-1 Population	a	Accuracy	5 Billion	5 Billion	5 Billion	5 Billion	5 Billion
			b	Bench Time	1.8 minutes	1.5 minutes	1.3 minutes	.9 minutes	1.3 minutes
2	Initial Performance	Benchmark-2 Basic needs	a	Accuracy	food, water shelter, med care	food, water shelter, med care	food, water shelter, med care	food, water shelter, med care	food, water shelter, med care
			b	Bench Time	1.9 minutes	1.6 minutes	1.4 minutes	1 minute	2.4 minutes
3	Initial Performance	Benchmark -3 Physical environment	a	Accuracy	collect resources mining agriculture	collect resources agriculture	collect, resources mining forestry	collect resources mining agriculture	collect resources mining foresty
			b	Bench Time	1.9 minutes	1.6 minutes	1.4 minutes	1.3 minutes	2.8 minutes
4	Initial Performance	Benchmark -4 Environmental interactions	a	Accuracy	erosion clearing vegetation siltation of streams	erosion overgrazing siltation of streams	erosion clearing vegetation siltation of streams	erosion clearing vegetation siltation of streams	erosion clearing vegetation siltation of streams
			b	Bench Time	2 minutes	1.7 minutes	1.4 minutes	.5 minutes	1.6 minutes
5	Initial Performance	Benchmark -5 Population pyramid	a	Accuracy	Horizontal bar = proportion of people in the cohort	Horizontal bar= estimated pop and the dates	Horizontal bar= proportion of people in the cohort	Horizontal bar= proportion of people in the cohort	Horizontal bar = proprotion of people in the cohort
			b	Bench Time	6 minutes	5.1 minutes	4.3 minutes	1 minute	1.8 minutes
6	Initial Performance	Benchmark -6 Stages 2 &3	a	Accuracy	birth rate = decl death rate = decl total pop = incr	birth rate = decl death rate = decl total pop = incr	birth rate = decl death rate = decl total pop = incr	birth rate = decl death rate = decl total pop = incr	birthrate = decl deathrate = decl total pop = incr
			b	Bench Time	2.3 minutes	1.9 minutes	1.6 minutes	1 minute	3.3 minutes
7	Initial Performance	Benchmark -7 Stage 4	a	Accuracy	birth rate = decl death rate = same total pop = incr	birth rate = low death rate = steady total pop = high	birth rate = decl death rate = decl total pop = incr	birth rate = decl death rate = stay same total pop = incr	birthrate decl deathrate = same total pop = incr
			b	Bench Time	.8 minutes	.6 minutes	.5 minutes	.4 minutes	1.6 minutes
8	Initial Performance	Benchmark -8 Low life expectancies	a	Accuracy	Africa	Europe	Africa	Africa	Africa
			b	Bench Time	3.3 minutes	2.7 minutes	2.3 minutes	1.5 minutes	4.7 minutes
9	Initial Performance	Benchmark -9 Infant mortality rate	a	Accuracy	Western Sahara	Africa	Western Sahara	Western Sahara	Western Sahara
			b	Bench Time	1 minute	.8 minutes	.7 minutes	.2 minutes	.7 minutes
10	Initial Performance	Benchmark-10 Net migration	a	Accuracy	+900	+900	+900	+900	+900
			b	Bench Time	.8 minutes	.6 minutes	.5 minutes	.4 minutes	.8 minutes

5.1.2 Usability Specification for MigModel

Table 2: MigModel Usability

Trial	Usability Attribute	Measuring Instrument		Value to be Measured	Current Level	Worst Judged Acceptable Levell	Planned Target Level	Best Target Level	Observed in Study
1	Initial Performance	B: Rural variable state	a	Accuracy	Far rt:-0.028 Far lt: 0.031	Far rt:-0.028 Far lt: 0.031	Far rt:-0.028 Far lt:-0.031	Far rt:-0.028 Far lt: 0.031	Far rt:-0.028 Far lt:-0.031
			b	Bench Time	5.5 minutes	4.7 minutes	4 minutes	1 minute	7 minutes
2	Initial Performance	B: Rural highest state match	a	Accuracy	Pbv: 0.02% match: 0.031	Pbv: 17.4% match: 0.0	Pbv: 1.6% match: 0.031	Pbv: 0.0% match: 0.031	Pbv: 26.8% match: 0.03
			b	Bench Time	4 minutes	3.5 minutes	2.9 minutes	1.5 minutes	3.4 minutes
3	Initial Performance	B: Rural variable county	a	Accuracy	Far rt:-0.108 Far lt: 0.112	Far rt:-0.108 Far lt: 0.112	Far rt:-0.108 Far lt: 0.112	Far rt:-0.108 Far lt: 0.112	Far rt:-0.108 Far lt: 0.112
			b	Bench Time	5 minutes	4.3 minutes	3.6 minutes	2 minutes	4 minutes
4	Initial Performance	B: Rural highest county match	a	Accuracy	Pbv: 0.0% match: 0.112	Pbv: 23.7% match: 0.106	Pbv: 0.0% match: 0.112	Pbv: 1.2% match: 0.113	Pbv: 6.2% match: 0.110
			b	Bench Time	2.5 minutes	2.1 minutes	1.8 minutes	1.2 minutes	7 minutes
5	Initial Performance	B: Distance to Coast state	a	Accuracy	Far Pbv: 824mi match: -0.099 Near Pbv: 8 mi match: 0.104	Far Pbv:711mi match: -0.099 Near Pbv:16mi match: 0.104	Far Pbv:731mi match: -0.099 Near Pbv:8 mi match: 0.104	Far Pbv:731mi match: -0.099 Near Pbv:8 mi match: 0.104	Far Pbv:731mi match: -0.099 Near Pbv:8 mi match: 0.104
			b	Bench Time	2 minutes	1.7 minutes	1.4 minutes	.5 minutes	3.6 minutes
6	Initial Performance	B: Distance to Coast county	a	Accuracy	Far Pbv: 820mi match: -0.341 Near Pbv: 5 mi match: 0.341	Far Pbv:820mi match: -0.341 Near Pbv: 5 mi match: 0.341	Far Pbv:820mi match: -0.341 Near Pbv: 5 mi match: 0.341	Far Pbv: 820mi match: -0.341 Near Pbv: 5 mi match: 0.341	Far Pbv: 820mi match: -0.341 Near Pbv: 5 mi match: 0.341
			b	Bench Time	3 minutes	2.6 minutes	2.1 minutes	.5 minutes	2.3 minutes
7	Initial Performance	B: Urban county	a	Accuracy	Most Urb Pbv: 96 match: 0.112 Least Urb Pbv: 0.0 match: -0.108 High Urb Pbv: 96 match:-0.112	Most Urb Pbv: 99 match: 0.112 Least Urb Pbv: 0.0 match: -0.108 High Urb Pbv: 60 match: 0.088	Most Urb Pbv: 100 match: 0.112 Least Urb Pbv: 0.0 match: -0.108 High Urb Pbv: 90 match: 0.112	Most Urb Pbv: 100 match: 0.112 Least Urb Pbv: 0.0 match: -0.108 High Urb Pbv: 97 match: 0.112	Most Urb Pbv: 100 match: 0.112 Least Urb Pbv: 0.0 match: -0.108 High Urb Pbv: 86 match:0.106
			b	Bench Time	4.5 minutes	3.8 minutes	3.2 minutes	2 minutes	4.5 minutes
8	Initial Performance	B: Urban and Distance to Coast county	a	Accuracy	Dist to Coast Pbv: 5 Urban Pbv: 100% match: 0.280	Dist to Coast Pbv: 8 Urban Pbv: 100% match: 0.280	Dist to Coast Pbv: 5 Urban Pbv: 100% match: 0.280	Dist to Coast Pbv: 5 Urban Pbv: 100% match: 0.280	Dist to Coast Pbv: 5 Urban Pbv: 100% match: 0.280
			b	Bench Time	4.3 minutes	3.6 minutes	3.1 minutes	2 minutes	2.8 minutes
9	Initial Performance	B: Importance Bar	a	Accuracy	match: 0.328	match: 0.328	match: 0.328	match: 0.328	match: 0.328
			b	Bench Time	1 minute	.9 minutes	.7 minutes	.25 minutes	1.5 minutes

5.1.3 Humpop Subject Results

Table 3: Humpop Subject Results

Trial	Measuring Instrument	Pilot 1	Pilot 2	Hum 1	Hum 2	Hum 3	Hum 4	Hum 5
1	B: population	5 billion	5 billion	5 billion	5 billion	5 billion	5 billion	5 billion
		1.5 minutes	2 minutes	1 minute	1.5 minutes	1.5 minutes	1.5 minutes	1 minute
2	B:basic needs	food, water shelter, med care	food, water shelter, med care	food, water shelter, med care	food, water shelter, med care	food, water shelter, med care	food, water shelter, med care	food, water shelter, med care
		1.5 minutes	2.3 minutes	2 minutes	2.5 minutes	2 minutes	2 minutes	3.5 minutes
3	B:physical environment	collect resource	collect resources mining agriculture	collect resources	forestry mining agriculture	collect resources argriculture	collect resources argriculture forestry	collect resources argriculture forestry
		2 minutes	1.7 minutes	5 minutes	2.5 seconds	1.5 minutes	2 minutes	3 minutes
4	B:environment interactions	erosion overgrazing pollution	erosion pollution overgrazing	erosion overgrazing pollution	erosion clearing veg siltation of streams	erosion clearing of land siltation of streams	erosion clearing of land siltation of streams	erosion overgrazing siltation of streams
		1.5 minutes	2.5 minutes	1.5 minutes	.5 minutes	1 minute	3 minutes	2 minutes
5	B:population pyramid	Horizontal bar= proportion of all people in the cohort	Horizontal bar= population of all people in the cohort	Horizontal bar= proportion of all people in the cohort	Horizontal bar = proportion of all people in the cohort	Horizontal bar = proportion of all people in the cohort	estimated pop in millions and the dates	horizontal bars = proportion of all people in that cohort
		7 minutes	5 minutes	2 minutes	2 minutes	2 minutes	1.5 minutes	1.5 minutes
6	B:stages 2 & 3	birth rate = decl death rate =decl total pop = incr	birth rate = decl death rate =decl total pop = incr	birth rate = decl death rate =decl total pop = incr	birth rate = inc death rate = decl total pop = incr	birth rate = decl death rate = decl total pop = incr	Not enough information	birthrate = decl death rate =decl total pop = incr
		2.5 minutes	2 minutes	3.5 minutes	3 minutes	1.5 minutes	4 minutes	4.5 minutes
7	B:stage 4	birth rate = decl death rate = same total pop = incr	birth rate = decl death rate = same total pop = incr	birth rate = decl death rate = same total pop = incr	birth rate = low death rate = steady total pop = high	birth rate = steady death rate = steady total pop = steady	not enough information	birthrate = decl death rate = steady total pop = decl
		.5 minutes	1 minute	3.5 minutes	1 minute	.5 minutes	1 minute	2 minutes
8	B:low life expectancies	Africa	Africa	Africa	Africa	Europe	Africa	Chad
		3 minutes	3.5 minutes	10 minutes	3.5 minutes	5 minutes	3 minutes	2 minutes
9	B:infant mortality rate	Western Sahara	Western Sahara	Western Sahara	Africa	Western Sahara	Western Sahara	Western Sahara
		1 minute	1 minute	1 minute	.25 minutes	.25 minutes	1.5 minutes	.5 minutes
10	B:net migration	+900	+900	+900	+900	+900	+900	+900
		.5 minutes	1 minute	1.5 minutes	.5 minutes	.5 minutes	1 minute	.5 minutes

5.1.4 MigModel Subject Results

Table 4: MigModel Subject Results

Trial	Measuring Instrument	Pilot 1	Pilot 2	Mig 1	Mig 2	Mig 3	Mig 4	Mig 5
1	B: Rural variable state	Far rt match: 0.096 Far lt match: 0.111	Far rt match: 0.028 Far lt match: 0.031	Far rt match: 50.2 Far lt match: 01.2	Far rt match:0.028 Far lt match: 0.031	Far rt match:0.028 Far lt match: 0.031	Far rt match:0.028 Far lt match:0.031	Far rt match: 0.028 Far lt match: 0.031
		4.5 minutes	6.5 minutes	6 minutes	11 minutes	4 minutes	5 minutes	8.5 minutes
2	B:Rural highest state match	Pbv: 11.5% Match: 0.019	Pbv: 0.04% Match: 0.031	Pbv: 98.0% Match: -0.031	Pbv: 0.0% Match: 0.031	Pbv: 17.4% Match: -0.229	Pbv: 1.6 % Match: 0.031	Pbv: 17.4 % Match: 0.000
		4.15 minutes	4 minutes	3 minutes	4 minutes	4 minutes	2 minutes	4 minutes
3	B:Rural variable county	Far rt match: 0.040 Far lt match: 0.112	Far rt match: 0.108 Far lt match: 0.112	Far rt match: 0.108 Far lt match: 0.112	Far rt match: 0.108 Far lt match: 0.112	Far rt match: 0.108 Far lt match: 0.112	Far rt match: 0.108 Far lt match: 0.112	Far rt match: 0.108 Far lt match: 0.112
		5.5 minutes	4.5 minutes	3.5 minutes	4.5 minutes	3 minutes	3 minutes	6 minutes
4	B:Rural highest county match	Pbv: 0.0 % Match: 0.112	Pbv: 0.0 % Match: 0.112	Pbv: 58.6 % Match: 0.01	Pbv: 0.0 % Match: 0.112	Pbv: 71.9 % Match: -0.014	Pbv: .8 % Match: 0.113	Pbv: .23.7 % Match: 0.106
		1 minute	4 minutes	4 minutes	5.5 minutes	4 minutes	4 minutes	15 minutes
5	B:Distance to Coast	Far Pbv: 731.1 Match: -0.099 Near Pbv: 8 Match: 0.104	Far Pbv: 734.0 Match: -0.099 Near Pbv: 8 Match: 0.104	Far Pbv: 416.1 Match: -0.305 Near Pbv: 5.0 Match: 0.341	Far Pbv: 731.1 Match: -0.099 Near Pbv: 8.0 Match: 0.104	Far Pbv: 711.0 Match: -0.099 Near Pbv: 16.6 Match: 0.104	Far Pbv: 731.1 Match: -0.099 Near Pbv: 8.0 Match: 0.104	Far Pbv: 731.1 Match: -0.099 Near Pbv: 8.0 Match: 0.104
		1 minute	3 minutes	3.5 minutes	4 minutes	4 minutes	3.5 minutes	3 minutes
6	B:Distance to Coast county	Far Pbv: 814.3 Match: -0.341 Near Pbv: 5 Match: 0.341	Far Pbv: 824.0 Match: -0.341 Near Pbv: 5 Match: 0.341	Far Pbv: 817.5 Match: -.341 Near Pbv: 814.3 Match: -0.340	Far Pbv: 295.0 Match: -0.184 Near Pbv: 31.0 Match: 0.105	Far Pbv: 731.1 Match: -0.099 Near Pbv: 16.6 Match: 0.104	Far Pbv: 820.8 Match: -0.341 Near Pbv: 5 Match: 0.341	Far Pbv: 824.0 Match: -0.341 Near Pbv: 5 Match: 0.341
		3 minutes	3 minutes	2.5 minutes	3 minutes	1 minute	3 minutes	2 minutes
7	B:Urban county	Most Urb Pbv: 0.0 Match: -.281 Least UrbPbv: 99.6 Match: -0.124 High UrbPbv: 50.2 Match: -.112	Most Urb Pbv: 100 Match 0.112 Least Urb Pbv: 0.0 Match: -0.108 High Urb Pbv 100 Match: 0.112	Most Urb Pbv: 97 Match: 0.112 Least Urb Pbv:0.01 Match: -0.108 High UrbPbv: 50.1 Match: 50.2	Most Urb Pbv:50.2 Match: -0.131 Least Urb Pbv: 0.0 Match: -0.028 High Urb Pbv: 100 Match: 0.031	Most Urb Pbv:99.6 Match: 0.031 Least Urb Pbv:-0.0 Match: -0.028 High Urb Pbv:99.6 Match: 0.03	Most Urb Pbv:99.6 Match: 0.112 Least Urb Pbv: 0.0 Match: -0.108 High Urb Pbv:97.6 Match: 0.113	Most Urb Pbv: 100 Match: 0.112 Least Urb Pbv: 0.0 Match: -0.108 High Urb Pbv:58.9 Match: 0.088
		4 minutes	5 minutes	1 minute	5 minutes	2.5 minutes	7 minutes	7 minutes
8	B:Urban and Distance to Coast county	Dist to coast Pbv: 5 Urban Pbv: 0.0 % Match: 0.124	Dist to coast Pbv: 5 Urban Pbv: 100% Match: 0.280	Dist to coast Pbv: Pbv: 416 Urban Pbv: 50 % Match: -.108	Dist to coast Pbv: 5 Urban Pbv: 100 % Match: 0.280	Dist to coast Pbv: 8 Urban Pbv: 100 % Match: 0.080	Dist to coast Pbv: Pbv: 5 Urban Pbv: 100% Match: 0.280	Dist to coast Pbv: 5 Urban Pbv: 100 % Match: 0.280
		4.5 minutes	4 minutes	2 minutes	4.5 minutes	2 minutes	2.5 minutes	3 minutes
9	B:Importance Bar	Match: 0.240	Match: 0.328	Match: -0.114	Match: 0.328	Match: 0.328	Match: 0.328	Match: 0.328
		1 minute	1 minute	1 minute	1 minute	2 minutes	1.5 minutes	2 minutes

5.1.5 Analysis of Usability Specification

Humpop

Overall the measure of learning for each benchmark task was very good compared with the expected Planned Level. The answers for the Current Level, Worst Acceptable Level, Planned Target Level, and Best Target Level responses were the same in most cases. For example, benchmark task two asked the user to identify four basic needs and all the responses for the different levels contained the same four basic needs. A reason for this might be that most users interpreted the information in the interface in the same fashion. There is a wide range in time completion of benchmark tasks. A reason for might be due to the different user experience levels.

MigModel

In MigModel the overall accuracy for each benchmark task was very good compared with the expected Planned Level. The accuracy of each benchmark task was affected by the use of the sliders in the interface. For example, benchmark task four has a range of 23% for the Predicted Best Value. It was at the discretion of the user where to place the sliders and run the models of simulation. The Worst Acceptable Level had subtle differences from the Planned Target Level relating to accuracy. For example, benchmark tasks five and eight has subtle differences between the Worst Acceptable Level and the Planned Target Level. The reason for this is due to how each participant interpreted the question and where the sliders were positioned for each benchmark task.

There was also a wide variance in time completion of the benchmark tasks. This might be explained by the participants' difficulty with placing the sliders. Further, when it came to finding the highest match value, there were several different approaches taken by the participants. Another factor that added to the time completion for each benchmark task was the tolerance on behalf of the users. Some users were more tolerant than others

and were more willing to stay engaged in completing the task while others quickly moved on. The users were not sure when they had found the highest match value, because it was not always a positive result.

5.2 QUALITATIVE RESULTS

The qualitative results were recorded from a questionnaire, Appendix C, that was given verbally after each participant completed the benchmark tasks. The responses were captured using a video camera. The users and the evaluator started a dialog session that allowed users to speak freely about their comments and concerns regarding the interface design. It was my impression that users would contribute more information verbally than in written form.

5.2.1 Humpop

Table 5: Humpop Verbal Protocol

Pilot 1	Felt lost when they came to the list of vocabulary words Felt that the vocabulary pages were without any direction The arrows leading from the bubbles were not very helpful, because after engaging in one bubble you had no way of knowing what was in the other bubbles without coming all the way out. Wished that there was a button that would allow them to go back to the previous steps, so when they felt lost they could trace their steps. The pictures were slow, didn't feel that they contributed significantly. The suggested sequence of study confused the user. Would have liked more directions.
Hum 1	Felt lost in the beginning. Thought it was really easy to work with. In the beginning when they missed a question they didn't know how to go back and answer the question. Wished they could have gone back a couple of screens.
Hum 2	Felt lost when they didn't know where to look for the answer. Operations in the systems were easy. Having a fast forward option that would speed up the movement of the pictures.

Table 5: Humpop Verbal Protocol

Hum 3	Felt lost when they didn't know exactly where to look. The system was pretty easy. The pictures annoyed them, they wanted to speedup and skip some sections in the first part.
Hum 4	Felt lost because they tried to follow the suggest sequence of study. Had a problem with finding everything. Thought all the answers were in one bubble. More diverse pictures, besides Kenya and Finland. Options -- a speedup function that would jump to the next section.
Hum 5	Never felt lost. The operations were easy. The pictures were slow, if you are not interested in the pictures then you should be able to skip through the pictures. A tree under the vocabulary to help with the flow -- indicate which vocabulary had charts, which were just vocabulary definitions.

5.2.2 MigModel

Table 6: MigModel Verbal Protocol

Pilot 1	Did not understand what mean value meant. Said that at first they confused Predicted best value with match value. Questions were very clear on exactly what was expected. Didn't know if they were to compare the two maps. Operations seemed to be easy. Annoyed by the time it takes to get an answer with county level analysis.
Pilot 2	Mouse was slow. Thought that the location of the red tic represented the Predicted best value. Operations seemed easy. County level of analysis took longer than the state level. Reluctant to find the highest match value because it took so long.
Pilot 3	In the beginning felt lost because they had no idea of the topic of migration. The operations seemed easy after playing around with the system. The system was slow when county level was being performed. State level was very fast. Thinks that the interface is very simple.
Mig 1	Felt lost at the beginning because, didn't read the directions. The operations were easy because of the highlighting on the screen. Said it was user-friendly. Nothing annoyed them. Didn't like the colors; wanted to change the colors because they were flat. Felt that it would be an aid to geography students.

Table 6: MigModel Verbal Protocol

Mig 2	Felt lost at the beginning because they were not used to the system. The operations seemed simple. Took long to load the county information. Would change the gray background because it was hard to see the white line marker. The white marker on a gray background didn't really stand out, maybe with a blue background. There was confusion about the red tic (mean value) thought that it related to the Predicted best value and said that maybe it should be up at the top. Add a description of what the colors mean in the legend, maybe up at the top somewhere.
Mig 3	Felt lost at the beginning, wasn't sure if they were doing the right thing. The layout of the interface was very simple; selecting the variables. The terms were a bit misleading, but the actual mechanics of it were very simple and straight forward. Would create a box that would give a demonstration of how to use the system, and better at explaining what the terms meant. For example something indicating how to find the predicted best values and match values on the screen. The demo could be up in the help box.
Mig 4	Didn't feel lost during the session. Everything is laid out just fine, everything was pretty easy. The county level was very slow. Nothing annoyed them abut the system other than the county level of analysis. It was very good to have two maps to compare simulations. Sometimes it was hard to click to get a 100% on the bar chart. A suggestion would be to have some arrows at the bottom that would help indicate when you reached 100%.

Table 6: MigModel Verbal Protocol

Mig 5	Felt lost trying to get the match value. They were confused in trying to get the lower map to match the upper map graphically. Had a hard time determining what was considered a high match value. Was annoyed when doing the counties.

5.2.3 Analysis of Verbal Protocol

Humpop

The participants for the Humpop tutorial felt lost in the beginning. The participants were not quite sure of what their first steps should be in trying to complete the benchmark tasks and the fact the no help options were available. Most of the participants agreed that the operations seemed easy. All navigation throughout the interface was done through a mouse. There were no keyboard entries of any kind and the user's options to choose different modules and module options were not complicated. There was a high number of requests for the ability to be able to backtrack through previous screens and to skip through the pictures that were located in the "Population Issues" unit. This was probably one of the most frustrating aspects of the interface. Once a user selected "Population Issues" unit they were locked into sitting through the entire display of pictures which lasts over a minute. A majority of the users would have liked to have more guidance in the interface when it came to finding information and getting started. The arrows which suggested the sequence to follow were not very helpful. It depended upon where the information was that the student was seeking. As the students progressed through the units, they did appear to become more comfortable with the navigation.

MigModel

The users experienced a lot of confusion in interpreting the terminology in the interface. For example, nowhere in the interface are the terms Predicted Best Value, Importance, and Match value clearly explained. A majority of the participants felt lost in the beginning for various reasons. One reason was that they had no way of knowing if they were performing the correct actions. For example, several benchmark tasks asked the users to find the highest match value. They had to interpret what was meant by

"high" match value. Further, unfamiliarity with the system also caused confusion. Being exposed to a new computing environment, many students felt very apprehensive toward exploring the system. Students agreed on many aspects. For example, most agreed that when the county level of analysis was selected the system took much longer to run the models than did the state level of analysis. There was total agreement that the operations in the interface were easy. This was due to having the only input coming from a mouse and not a keyboard. Also they agreed that highlighting was very good and they felt that they knew what the system was doing when they were executing commands.

5.3 QUANTITATIVE RESULTS

The quantitative questionnaire gave the users an opportunity to rate the system, user interface, and their overall experience [1]. The average rating given for each category is represented by bold and underline. The results for each category was taken from the average response from all the users.

5.3.1 Humpop

Post Experiment Questionnaire

Please circle the number that most accurately reflects your impression of using the system.

• OVERALL REACTIONS TO THE SOFTWARE

terrible wonderful

0 1 2 3 4 5 6 **7** 8 9 NA

difficult easy

0 1 2 3 4 5 6 7 **8** 9 NA

frustrating satisfying

0 1 2 3 4 5 6 **7** 8 9 NA

Inadequate power adequate power

0 1 2 3 4 5 6 **7** 8 9 NA

dull stimulating

0 1 2 3 4 5 6 **7** 8 9 NA

rigid flexible

0 1 2 3 4 **5** 6 7 8 9 NA

• SCREEN

Characters on the computer screen

hard to read easy to read

0 1 2 3 4 5 6 7 **8** 9 NA

Highlighting on the screen simplifies task

not at all very much

0 1 2 3 4 5 6 **7** 8 9 NA

Organization of information on screen

confusing very clear

0 1 2 3 4 5 6 **7** 8 9 NA

Sequence of screens

confusing very clear

0 1 2 3 4 5 6 7 **8** 9 NA

• TERMINOLOGY AND SYSTEM INFORMATION

Use of terms throughout system

inconsistent consistent

0 1 2 3 4 5 6 **7** 8 9 NA

Computer terminology is related to the task you are doing

never always

0 1 2 3 4 **5** 6 7 8 9 NA

Position of messages on screen

inconsistent consistent

0 1 2 3 4 5 6 **7** 8 9 NA

Messages on screen which prompt user for input

confusing clear

0 1 2 3 4 5 6 7 **8** 9 NA

Computer keeps you informed about what it is doing

never always

0 1 2 3 4 **5** 6 7 8 9 NA

Error messages

unhelpful helpful

0 1 2 3 4 5 6 7 8 9 **NA**

- LEARNING

Learning to operate the system

difficult easy

0 1 2 3 4 5 6 7 **8** 9 NA

Exploring new features by trial and error

difficult easy

0 1 2 3 4 5 6 7 **8** 9 NA

Remembering names and use of commands

difficult easy

0 1 2 3 4 5 6 7 **8** 9 NA

Tasks can be performed in a straight-forward manner

never always

0 1 2 3 4 5 6 **7** 8 9 NA

Help messages on the screen

unhelpful helpful

0 1 2 3 4 5 6 7 **8** 9 NA

Supplemental reference materials

confusing clear

0 1 2 3 4 5 6 7 **8** 9 NA

• SYSTEM CAPABILITIES

System speed

too slow fast enough

0 1 2 3 **4** 5 6 7 8 9 NA

System reliability

unreliable reliable

0 1 2 3 4 5 6 7 **8** 9 NA

System tends to be

noisy quiet

0 1 2 3 4 5 **6** 7 8 9 NA

Correcting your mistakes

difficult easy

0 1 2 3 4 5 6 **7** 8 9 NA

Experienced and inexperienced users' needs are taken into consideration

never always

0 1 2 3 4 5 6 **7** 8 9 NA

5.3.2 MigModel

Post Experiment Questionnaire

Please circle the number that most accurately reflects your impression of using the system.

• OVERALL REACTIONS TO THE SOFTWARE

terrible wonderful

0 1 2 3 4 5 **6** 7 8 9 NA

difficult easy

0 1 2 3 4 5 **6** 7 8 9 NA

frustrating satisfying

0 1 2 3 4 **5** 6 7 8 9 NA

Inadequate power adequate power

0 1 2 3 4 5 **6** 7 8 9 NA

dull stimulating

0 1 2 3 4 5 **6** 7 8 9 NA

rigid flexible

0 1 2 3 4 5 6 **7** 8 9 NA

• SCREEN

Characters on the computer screen

hard to read easy to read

0 1 2 3 4 5 6 7 **8** 9 NA

Highlighting on the screen simplifies task

not at all very much

0 1 2 3 4 5 6 **7** 8 9 NA

Organization of information on screen

confusing very clear

0 1 2 3 4 5 6 **7** 8 9 NA

Sequence of screens

confusing very clear

0 1 2 3 4 5 6 7 **8** 9 NA

- TERMINOLOGY AND SYSTEM INFORMATION

Use of terms throughout system

inconsistent consistent

0 1 2 3 4 5 6 7 **8** 9 NA

Computer terminology is related to the task you are doing

never always

0 1 2 3 4 5 **6** 7 8 9 NA

Position of messages on screen

inconsistent consistent

0 1 2 3 4 5 6 7 **8** 9 NA

Messages on screen which prompt user for input

confusing clear

0 1 2 3 4 5 6 7 **8** 9 NA

Computer keeps you informed about what it is doing

never always

0 1 2 3 4 5 6 7 **8** 9 NA

Error messages

unhelpful helpful

0 1 2 3 4 5 6 7 8 9 **NA**

• LEARNING

Learning to operate the system

difficult easy

0 1 2 3 4 5 **6** 7 8 9 NA

Exploring new features by trial and error

difficult easy

0 1 2 3 4 5 6 7 **8** 9 NA

Remembering names and use of commands

difficult easy

0 1 2 3 4 5 **6** 7 8 9 NA

Tasks can be performed in a straight-forward manner

never always

0 1 2 3 4 5 6 **7** 8 9 NA

Help messages on the screen

unhelpful helpful

0 1 2 3 4 5 6 7 8 9 **NA**

Supplemental reference materials

confusing clear

0 1 2 3 4 **5** 6 7 8 9 NA

• SYSTEM CAPABILITIES

System speed

too slow fast enough

0 1 2 3 **4** 5 6 7 8 9 NA

System reliability

unreliable reliable

0 1 2 3 4 5 6 7 8 **9** NA

System tends to be

noisy quiet

0 1 2 3 4 5 6 7 **8** 9 NA

Correcting your mistakes

difficult easy

0 1 2 3 4 5 6 7 **8** 9 NA

Experienced and inexperienced users' needs are taken into consideration

never always

0 1 2 3 4 5 6 7 **8** 9 NA

5.3.3 Analysis of User Ratings

Humpop

As overall reactions toward the software, the users gave good ratings because the operations were simple and the interface did not frustrate the users. The screens in the interface received good ratings because the screens were easy to read and the interface layout was well highlighted and very spacious. As a result the users felt that the screens did not appear clustered or distracting. The terminology received a lower rating because much of the material consisted of vocabulary words unfamiliar to the user. System feedback received a low rating. The interface did not do a good job in informing the users as to where they were in the system and what would be the consequences of their actions. The categories related to learning received high marks. The users felt that learning the operations was straight-forward and easy. The system capabilities received a low rating regarding speed. This was due to the sequence of pictures. The participants wanted a way to either exit the pictures unit or speed up the picture displays. The other areas relating to system capabilities received good marks. It took users awhile to get familiarized with the user interface.

MigModel

The overall reactions to the software were not high. This was due to the fact that users really were not sure if they were taking the right steps in completing the benchmark tasks. Another factor related to the speed of county level of analysis. The screen, terminology, and system information received high marks because the screen layout in MigModel is very good. The characters are easy to read and the screen layout is very clear, not cluttered or distracting. The interface does a good job in keeping the user informed of what actions are currently being taken by the system. This gives good system feedback to the user. The initial response in learning the system did not receive

high ratings. Again this is probably due to the fact that students felt very unsure of their actions. Students were probably trying to interpret the terminology from the questions and the interface. In regards to the system capability, the speed received low marks. The is probably a reflection of the slow loading of the county level of analysis. High marks were given in the other areas relating to system capabilities. This is probably a result from the interface being menu driven and mouse operated. These two characteristics prevented the users from seeing any error messages.

6.0 RECOMMENDATIONS

After analyzing the data and the users' comments, I conclude that particular areas need some fine tuning. I have created a list of suggestions that should improve the overall performance and enhance the displays for both MigModel simulation software and the Humpop tutorial.

6.1 Humpop

A. **Suggestion:** Changing terms

Description: - The term "continue" is used to return back to the menu selection as well as proceed to the next page of information. The term "continue" should be replaced with "return" when the following action returns to a menu selection. See Table 7.

Table 7: General Vocabulary Population Studies

Issue: **Vocabulary Words**	Suggestion
Birth Control	Replace the term "continue" with "return"
Carrying capacity	same
Census	same
COHORT	same
Demographic Transition	same
Demography	same
Migration	same
Natural change	same
Overpopulation	same
Population geography	same
Underpopulation	same
POPULATION PROJECTION	Replace the last term "continue" with "return"
POPULATION PYRAMID	same

B. **Suggestion:** Consistent headers

Description: - There should be consistency with the headers in the individual modules. For example, in the Population Studies module; the header is "General Vocabulary" and the header should be "Population Studies Vocabulary" which is consistent with the other module headers.

C. **Suggestion:** Returning to the root menu

Description: - The user should be able to return to the root menu and not have to go through several screens before given the option to return. See Table 8.

Table 8: Stages of Demographic Transition

Issue: **General**	Suggestion
Need to be able to return to root menu from every page	Add a return option along with the continue option
Last page	Change the term "continue" to "return"

D. **Suggestion:** Begin Simulation button

Description: - The actions of the Begin Simulation button are not explained anywhere in the interface. There seems to be some confusion between the commands "click on topic of interest" and "Begin Simulation". The user is confused upon which action to take not realizing that "Begin Simulation" takes them to another program. A suggestion is to change the wording for the button "Begin Simulation" to "Go to Intlpop Program". This would be more helpful because this option takes the user into a new program that is totally different from Humpop.

E. **Suggestion:** Help Balloons

Description: - Help balloons could give more description to the command buttons located throughout the interface. For example the return's should indicate where you are returning to. (i.e. to the "Population Studies Vocabulary" menu or to the main menu) Other buttons that would gain from more description are the "quit" and the "Go to Intlpop Program" buttons.

F. **Suggestion:** History button.

Description: - A history button could be added to the interface to let the users know which unit they have just visited and in what order they were visited. See Fig 5.

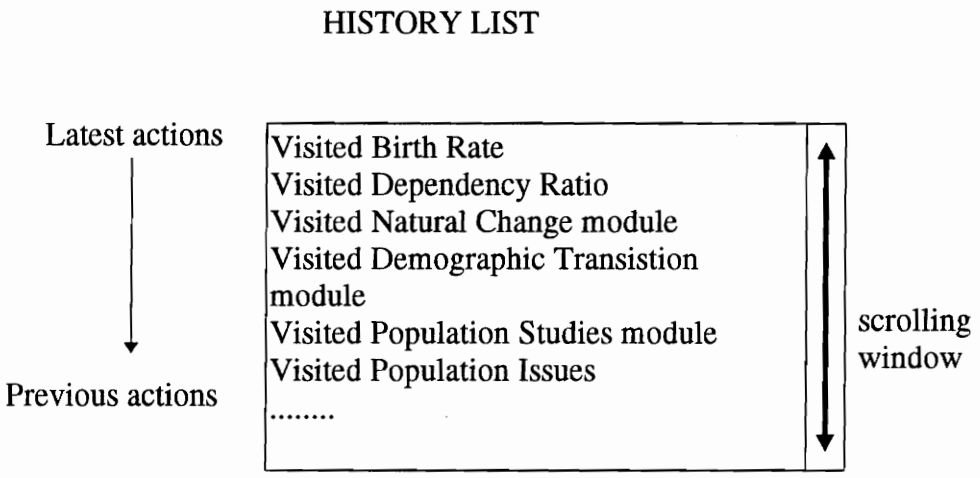

Figure 5: History button

G. **Suggestion:** Breaking a sequence of pictures

Description: The user should not be forced to sit through an entire display of pictures. In the "Population Issues" module there are no options to return to the main menu once the unit has been started. A "return" button could appear along side of the already existing "continue" button. This would give the user the ability to break a long series of pictures.

H. **Suggestion:** <u>You are Here button</u>

Description: - The "You are Here" button could give a graphical representation of where the user is in the system. The users were unsure of their location once they entered a unit. They were forced to return to the main menu to get a sense of where they were in the system. See Fig 6.

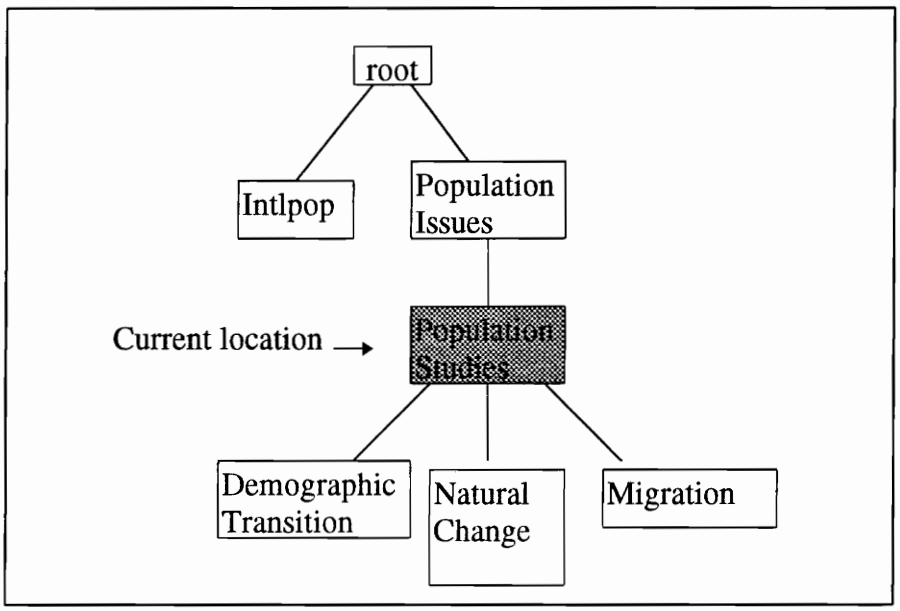

Figure 6: You Are Here button

6.2 MigModel

A. **Suggestion:** Walk-through tutorial

Description: - Under the Help section "How to build a model" the text should be replaced with an actual model being built. All the participants were in agreement that if a walk-through model was available it would help them understand Predicted Best Value, mean (red tic), Importance, Match, and sliders.

B. **Suggestion:** Definition page

Description: - Have one page located up in the help menu dedicated solely to define Predicted Best Value, mean (red tic), Importance, and Match. Users who are trying to define terminology should only have to go to one section. A definition page would serve this purpose.

C. **Suggestion:** Highlighting

Description: - The wording Predicted Best Value and Importance are grayed out in the interface, giving an indication that they are not options which can be selected by the user. Instead Predicted Best Value and Importance could appear in black bold letters indicating that they are currently active. If this happens then both sliders could appear in white. This suggestion would meet the guidelines for consistency.

D. **Suggestion:** Background color

Description: - Pink is a not a good background color, gray would be a better selection. Pink is a very distracting color while gray is a more subdued tone.

E. **Suggestion:** Map Color Definitions

Description: - Under the help section "Map Color Definitions" the word bright cyan should be replaced with greenish blue. More people can recognize the term greenish blue as opposed to the term "cyan." Also in the help section, under "Map Color Definitions," the word state should be added with county. For example it should read county or state. At present, the only level of analysis is county even though there are two levels of analysis -- state and county.

F. **Suggestion:** Select Variables

Description: - Under the Select Variables menu there is not a description of what each variable is measuring. A help balloon could be added to indicate what each variable is measuring. When a user highlights a variable, a small balloon could appear giving a information about what it measures.

7.0 Further Evaluation

Formative evaluation primarily addresses issues pertaining to the early development of an interface and redesign. This evaluation process has several steps that allow the developers and users to perform early testing of an interface as well as testing throughout the development stages of a project [2]. After formative evaluation has been performed the next steps are to:

- Examine at the quantitative and qualitative results
- Determine the importance of the problems
- Estimate the cost - resources (time, money) needed for each of the proposed solutions
- Make final decisions on how to address each problem
- Redesign

APPENDICES

A. Informed Consent Form

This form constitutes informed consent by you to participate in this study. Please read this entire form and then sign it if you agree to participate.

Thank you for participating in this research. This study is being conducted by Kevin L. Stringer, graduate student, Dr. Cliff Shaffer, and Dr. Rex Hartson, advisor, for the **Project GeoSim** research group. The study examines the usability of a software interface. If you choose to participate in this research, you will be asked to perform several exercises on a computer, followed by a written questionnaire, and then some verbal questions. No physical activities other than using a mouse and typing will be required. The experiment will last one hour. You will be paid $5.00. There are no known sources of discomfort in this study. Your participation in the experiment will greatly aid the research team in developing a quality user interface.

If at any time during the experiment you feel that you cannot participate further or that you need help understanding any portion of the task or equipment, please inform the experimenter.

Information collected during the study will be confidential. The experiment will be video taped. The video tapes will be confidential information reviewed by Kevin L. Stringer and will be erased by December 1, 1994. The tapes will be kept in a locked compartment with only Kevin L. Stringer having access to these tapes. The data derived from the video tapes will be used to collect qualitative results. The only other person who might see the information from this experiment is Ms. Susan Keenan, a Ph.D. student doing research in usability problems.

As a participant in the study, you have certain rights:

1. You may withdraw from the experiment at any time for any reason.

2. At the conclusion of your participation, you may see your data, if you desire. If you decide to withdraw your data, please inform the experimenter immediately.

If you have any questions about this experiment, please contact the researcher or his advisor:

- Researcher: Kevin L. Stringer, home: (703) 951 - 2489

- Advisor: Dr. Rex Hartson. office: 101 Femyor (703) 231 - 4857

Additional questions regarding your rights as a participant should be addressed to Dr. E. R. Stout, Chairman of the Institutional Review Board at (703) 231 - 9359 (306 Burruss Hall).

Your signature below indicates that you have read this document (Participant's Informed Consent Form) in its entirety, that your questions have been answered, and that you consent to participate in the study described.

Printed name: _____

School Address: _____

_____

_____

Student ID #: _____

Signature:_____ Date: _____

B. Nondisclosure Agreement Form

This form constitutes agreement by you not to discuss any part of this experiment with anyone other than the researcher, his advisor or member of the Institutional Review Board until September 30, 1994.

To protect the integrity of this study you are asked to agree not to discuss any part of this experiment or your participation in it with anyone other than the researcher, his advisor or members of the Institutional Review Board any time before September 30, 1994. After that time, you may discuss the experiment and your participation with anyone.

The purpose of this agreement is to ensure that other participants in the study are not influenced in any way by your experience in the experiment. Such influence could invalidate the results of the experiment.

If you have any questions about this experiment, please contact the researcher or his advisor:

- Researcher: Kevin L. Stringer, home: (703) 951 - 2489

- Advisor: Dr. Rex Hartson. office: 101 Femyor (703) 231 - 4857

Additional questions regarding your rights as a participant should be addressed to Dr. E. R. Stout, Chairman of the Institutional Review Board at (703) 231 - 6077 (301 Burruss Hall).

Your signature below indicates that you have read this document (Participant's Nondisclosure Agreement) in its entirety and that you agree not to discuss the

experiment with anyone other than the researcher, his advisor or the Institutional Review Board any time before September 30, 1994.

Printed name: _____

School Address: _____

_____

_____

Student ID #: _____

Signature:_____ Date: _____

C. Qualitative Questionnaire: Verbal Protocol

We **greatly appreciate** your willingness to work with us in evaluating this new teaching approach for introductory geography. As a final request, would you please answer the following questions regarding your experience using this simulation program.

1. Did you ever feel as if you did not know what to do or that you did not know what you were doing?

   If yes, please explain.

2. In general, what types of operations did you find easy?

3. What did you find annoying in the system?

4. What would you change if you were the designer?

5. Do you have any additional comments?

Thank you for your participation

D. Demographic and User Experience Questionnaire

Pre Experiment

Please answer the following questions. For each question, circle the answer that matches your response.

Age:_____

Number of years of college completed:_____

Gender: Male Female

Have you ever used a computer program operated by a mouse before?

a) no
b) yes, but only a few times
c) yes, many times

Have you ever used or seen someone use any Project GeoSim software, such as *IntlPop*?

Yes No

How much experience do you have using computers?

a) I have never, or almost never used a computer before.

b) I consider myself a novice computer user.

c) I consider myself an average computer user.

d) I consider myself an experienced computer user.

E. Subject Instruction for Humpop

Human Population Change

Proceed through the human population change tutorial at your own pace using the suggested sequence of study. You may repeat any topics of interest at any time. As you progress through the tutorial, answer the following questions.

Task 1

What is the current population of the world?

Answer:

Task 2

What are four basic needs that must be met in every human population?

Answer:

Task 3

Identify three ways in which human populations modify the physical environment.

Answer:

Task 4

As human population increases, human-physical environmental interactions intensify. Identify three specific consequences of these intensified interactions.

Answer:

Task 5

What do the bars on a population pyramid represent?

Answer:

Task 6

Developing countries are usually associated with stages 2 and 3 of the demographic transition. Describe what is happening with birth rates, death rates, and total population in those stages.

Answer:

Task 7

Developed countries are usually associated with stage 4 of the demographic transition. Describe what is happening with birth rates, death rates, and total population in that stage.

Answer:

Task 8

Which continent has the most countries with low life expectancies?

Answer:

Task 9

In 1990, which country had the highest infant mortality rate?

Answer:

Task 10

What is the net migration for country A in the net migration illustration.

Answer:

F. Subject Instruction for MigModel

MIGRATION IN THE UNITED STATES

Migration Modeling

In this exercise you are to experiment with a new migration modeling program from Project **GeoSim**. In this module you attempt to discover factors that might be related to migration in the United States at the county and state levels. You will be trying to explain the patterns of migration as shown on a map of the US by selecting variables and "Predicted Best Values" for those variables in building a model that predicts what actually happened.

DEFINITIONS and PROGRAM EXPLANATION:

The program has on-line help that will assist you when you first begin, but a few definitions are also needed:

1) PREDICTED BEST VALUE - represents that value that is most likely to induce immigration (migration to that place). For example, you might select snowfall as a variable. Setting a predicted best value at the left end of the slide bar (lowest snowfall) suggests that you think that people are pulled toward areas where there is little or no snowfall, and ALSO that they are pushed away from areas where there is high snowfall.

2) IMPORTANCE - when you are building a model with more than one explanatory variable, it is possible to make each variable more or less significant in the explanation. By default all variables selected have equal significance. You may move the slide bar from the far left (almost no significance as if the variable were not selected at all), to the far right (the variable is very dominant in the model).

3) MEAN VALUE - the **red tic** along each slide bar represents the US average for that variable. For instance, the slide bar for distance from coast for the states shows a mean value around 216 miles. This value is shown simply for your information in making your decisions on predicted best values.

4) MATCH - the match value tells you how closely two maps resemble each other. The first map (upper right corner) is the actual migration pattern for a chosen decade. The second map (lower right corner) is a reflection of the values chosen by the user. The more the two maps resemble one another the higher (+) the value. The less the two maps resemble one another the smaller (-) the value.

Task 1

Read the question carefully and be sure you understand it before you begin.

Write you answer in the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

Select the **Decade** of the 1950's.

Select the **State** level of analysis.

Click the Select Variables button, then select the variable **Rural** from the list and click DONE. A slide bar for setting the predicted best value appears. From the map pattern and your own knowledge of the states of the US, which end of the slide bar do you think will provide a better model? (i.e. Do you think people move to states that are very rural, or very urban?)

Give the match values after running the models.

Move the slide bar to the far right end and then click Run Model.

Match Value = _____

Move the slide bar to the far left end and then click Run Model.

Match Value = _____

Having run the models at both settings, briefly describe what these two models have told you about the area to which people migrated.

When you have answered the question, tell the experimenter that you are done.

Task 2

Read the question carefully and be sure you understand it before you begin.

Write you answer in the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

If the **Decade** of the 1950's is not selected then select it.

If the **State** level of analysis is not selected then select it.

What value of **Rural** provides the highest match value between the two maps?

Predicted best value = _____Match Value = _____

Note: This may take several attempts.

When you have answered the question, tell the experimenter that you are done.

Task 3

Read the question carefully and be sure you understand if before you begin.

Write you answer on the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

MigModel also allows you to study the migration patterns at the county level. Select County, and repeat the analysis. From your own knowledge of the counties of the US and the pattern on the actual migration map, which end of the slide bar do you think will provide a better model for the **Rural** variable? (i.e. Do you think people move to counties that are very rural, or very urban?)

Give the match values after running the models.

Move the slide bar to the far right end and then click Run Model.

Match Value = _____

Move the slide bar to the far left end and then click Run Model.

Match Value = _____

Having run the models at both settings, briefly describe what these two models have told you about the area to which people migrated. Is the result the same as that for the states?

When you have answered the question, tell the experimenter that you are done.

Task 4

Read the question carefully and be sure you understand it before you begin.

Write you answer in the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

If the **Decade** of the 1950's is not selected then select it.

If the **County** level of analysis is not selected then select it.

What value of **Rural** provides the highest match value between the two maps?

Predicted best value = _____Match Value = _____

Note: This may take several attempts.

When you have answered the question, tell the experimenter that you are done.

Task 5

Read the question carefully and be sure you understand it before you begin.

Write you answer in the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

Next, clear all variables and select the state level again. Then select **Distance to Coast.** Give the values for the hypothesis that people wanted to live as far from the coast as possible.

Predicted best value _____ Match Value _____

Give the values for the hypothesis that people wanted to live as close to the coast as possible.

Predicted best value _____ Match Value _____

Are the shores of the great lakes considered coasts in these models? How do you know?

When you have answered the question, tell the experimenter that you are done.

Task 6

Read the question carefully and be sure you understand it before you begin.

Write you answer in the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

Select the county analysis level for **Distance to Coast**. Set the predicted best value to hypothesize that people will be attracted to the coast. What match value did this model give?

Predicted best value _____ Match Value _____

Rerun the model with the hypothesis that people moved as far away from the coast as possible. What are the values now?

Predicted best value _____ match value _____

Did people tend to move toward or away from the coast?

TOWARD AWAY

When you have answered the question, tell the experimenter that you are done.

Task 7

Read the question carefully and be sure you understand it before you begin.

Write you answer in the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

While the **Distance to Coast** variable explains some of the pattern of migration, study the upper map. Note that in the Midwest, there are many counties far from the coast that are still in yellow (highest immigration). Perhaps those are urban areas. Clear all variables and then select **Urban**. First, run the model that hypothesizes that people moved to the most urban counties in the US.

Predicted best value _____ Match Value _____

Run another model to hypothesize that people move to the least urban locations in the US.

Predicted best value _____ Match Value _____

Find the predicted best value that provides highest match value for **Urban.**

Predicted best value _____ Match Value _____

When you have answered the question, tell the experimenter that you are done.

Task 8

Read the question carefully and be sure you understand it before you begin.

Write you answer in the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

If county level of analysis is not selected then select.

Select both **Urban** and **Distance to Coast** variables.

It is possible that people wanted to move to places either near the coast or to urban counties. Run the model hypothesizing that people will be attracted to the coast and that people moved to the most urban counties in the US.

Distance to Coast predicted best value _____

Urban predicted best value _____

Match Value _____

Did combining these two variables make a better model of the migration patterns in the 1950's ?

When you have answered the question, tell the experimenter that you are done.

Task 9

Read the question carefully and be sure you understand it before you begin.

Write you answer in the space provided.

When you understand the question and are ready to begin, tell the experimenter that you are ready.

When you have answered the question, tell the experimenter that you are done.

The county level of analysis and both **Distance to Coast** and **Urban** variables should still be selected. Adjust the importance of the two variables (do not change the predicted best values) so **Distance to Coast** is about twice as important as **Urban.** You may do this by sliding the Importance bar for **Distance to Coast** to the far right. Did that improve the match of your model?

Match Value _____

When you have answered the question, tell the experimenter that you are done.

REFERENCES

1. Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an Instrument Measuring User Satisfaction of the Human-Computing Interface. *Proceedings of CHI Conference on Human Factors in Computing Systems,* New York: ACM, 213-218.

2. Hartson, H. R., & Hix, D. (1993). DEVELOPING USER INTERFACES: *Ensuring Usability Through Product & Process* (pp.221-248,283-340). Toronto: Wiley.

3. Shneiderman, Ben. (1992). *Designing the User Interface: Strategies for Effective Human-Computer Interaction 2nd ed.* (pp.325-330). Reading, Massachusetts: Addison-Wesley Publishing Company.