
AC 2012-4422: USABILITY EVALUATION OF A PROBLEM SOLVING ENVIRONMENT FOR AUTOMATED SYSTEM INTEGRATION EDUCATION USING EYE-TRACKING

Punit Deotale, Texas A&M University

Dr. Sheng-Jen "Tony" Hsieh, Texas A&M University

Dr. Sheng-Jen ("Tony") Hsieh is a professor in the Dwight Look College of Engineering at Texas A&M University. He holds a joint appointment with the Department of Engineering Technology and the Department of Mechanical Engineering. His research interests include engineering education, cognitive task analysis, automation, robotics and control, intelligent manufacturing system design, and micro/nano manufacturing. He is also the Director of the Rockwell Automation Laboratory at Texas A&M University, a state-of-the-art facility for education and research in the areas of automation, control, and automated system integration.

Usability Evaluation of a Problem Solving Environment for Automated System Integration Education Using Eye-tracking

Abstract

Research suggests that realistic practice using authentic learning environments leads to better transfer of skills. Based upon input from industry engineers with expertise in designing and building automated systems, two problem solving environments (PSEs) were developed. Each PSE provides a virtual environment for building, testing, and validating designs for a process to be automated. The user is presented a toolbox containing equipment and building blocks of automation such as sensors and actuators. The PSEs support students in design problem-solving activities such as (1) reviewing the problem, (2) understanding the process to be automated, (3) line balancing, (4) layout, (5) simulation, and (6) cost analysis.

This paper presents preliminary results from a usability evaluation of one of the PSEs. The evaluation utilized a Facelab desktop mounted eye-tracking system. Users' eye movements were tracked using a camera and sensors to determine gaze direction and fixations. The data collected from the evaluation was used to determine if use of visual cues improved the usability of the PSE. Results suggest that use of visual cues for gaze direction improved the usability of the PSE application, based on faster task completion times and improved navigability.

Introduction

In earlier work, the authors described the design of a web-based problem-solving environment (PSE) for automated system integration education¹. For this paper, we describe preliminary results of a usability evaluation of this PSE, in which we investigated the effect of visual cues in helping users to navigate through the user interface.

Visual cues direct a user's gaze towards predetermined areas of interest on the screen. In the PSE, these visual cues help the user to follow the correct set of actions to achieve the desired goal. A large amount of data on the screen can be confusing. It can be useful to provide guidance in choosing the correct sequence of actions, such as a small rectangle or a red dot in the area that the user is supposed to click next.

In the user interface (UI) for the PSE, main menu buttons were placed on an easily accessible menu bar at the bottom of the page. Design elements were placed in a natural left to right reading order. For example, the first button that the users were supposed to click was placed on the extreme left while the last button in the navigational flow was placed on the extreme right of the screen.

Experimental Design

Three prototypes were developed, each with different levels of visual cues. Participants were given a specific task to complete using the PSE. The task was to select machines, adjust their Time and Cost parameters, and then connect them in a layout according to the guidelines given on the first page of the PSE. Upon task completion, participants answered questions based on their experience using the PSE.

Participants

Fifteen participants between the ages of twenty to thirty volunteered for the study. All of the participants were undergraduate students at a university in the southwestern U.S.

Apparatus

Seeing Machines Inc's Facelab eye-tracking system was used for the experiments. Facelab is a desktop mounted eye-tracking system, consisting of two cameras mounted on a stereo head and an infrared (IR) pod (Figure 1). The IR pod emits infrared light, which is reflected off users' eyes; the reflection is recorded by the cameras to track the eye movements.

A software package called Facelab 5.0, which comes bundled with the system, was used to record data. A software suite called Eyeworks from Eyetracking Inc. was used along with Facelab for data collection and analysis. The Eyeworks suite includes three software applications:

- *Eyeworks Design* is used to design custom scripts to be used in the experiments.
- *Eyeworks Record* records the data necessary for analysis.
- *Eyeworks Analyze* is an analysis tool that can be used to do visual analysis on the eye-tracking data and export statistical data out of the software.



Fig. 1. Facelab eye-tracking system



Fig. 2. Experiment in progress

Procedure

Fifteen participants were grouped randomly into three groups of five each:

- Prototype I: Participants in this group were presented with Prototype I, which had no visual cues. This group served as the control group.
- Prototype II: Participants in this group were presented with Prototype II, which used various colored rectangles as visual cues.
- Prototype III: Participants in this group were presented with Prototype III, in which red dots are presented as visual cues.

The experiments proceeded as follows:

- Participants were given an information sheet explaining the experiment, significance of the research and potential risks involved with the experiment.
- Participants were seated in front of a computer screen and the eye tracker was set up as shown in Figure 2.
- The eye tracker was configured for each participant before the experiment. Participants were asked to remain as steady as possible during the calibration.
- Information about the PSE was presented in the first step of the experiment.
- Participants were asked to use the PSE to fulfill a task described on the “Problem” screen (design of an automated system for cell phone assembly).
- Participants were asked to answer some simple questions after the end of the experiment.

Results

Eye-tracking records a large amount of data. It is assumed that a large number of fixations on a particular region indicate a significant area of interest. For this evaluation, three factors were important in determining the areas of visual interest:

- A large number of fixations indicates an important area of visual interest.
- Frequency of gaze is also an important factor in determining an area of visual interest. The more frequent the gaze, the more important the regions. In other words, if a particular region on the screen attracts the user’s attention, the number of gaze hits to that particular region is greater than that of any other region on the screen.
- Scan-paths give information about users’ eye movement patterns. Scan-paths were compared for each of the prototypes on the basis of time and effort taken by the user to find the significant areas of interest. An uncluttered and short scan path is taken to mean that the user was able to easily identify the significant areas on the screen, whereas a lengthy and cluttered scan path is taken to show that the user could not differentiate between the significant and insignificant areas on the screen.

Prototype I

Prototype I did not include any visual cues. Figure 3 shows a screenshot of the application window. Five users were presented with this prototype. The scanpath from one particular experiment, as shown in Figure 4, indicates how the user’s eyes moved about the screen while looking for relevant information. The green circle denotes the starting point and the red circle denotes the point where the user’s gaze ended. A cluttered scan path suggests that the user had low confidence that he/she was looking at the correct area.

Figure 5 shows a heat map pattern of the same experiment. The color red denotes the areas that attracted the user’s gaze the most and blue denotes moderate gaze activity. The heat map shows a high concentration of red areas on the left side of the screen indicating that the user’s eyes are attracted primarily to one part of the screen. The user overlooks or does not notice other areas of interest that are important in the navigational flow. For example, the areas labeled Time and Cost are significant areas of the PSE; but, as seen in the heat map, there are no red areas on the Time and Cost fields.

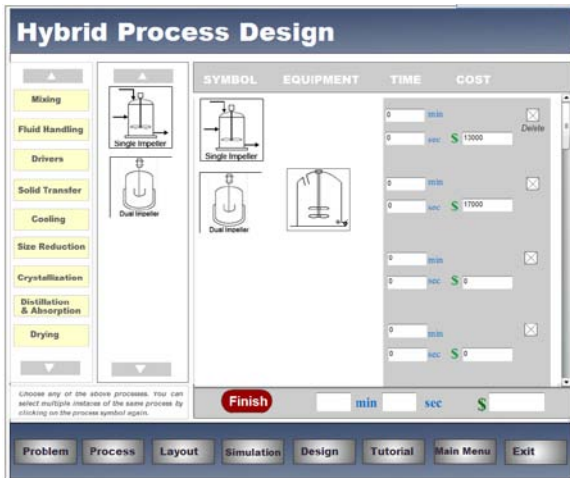


Fig. 3. Prototype I – No visual cues

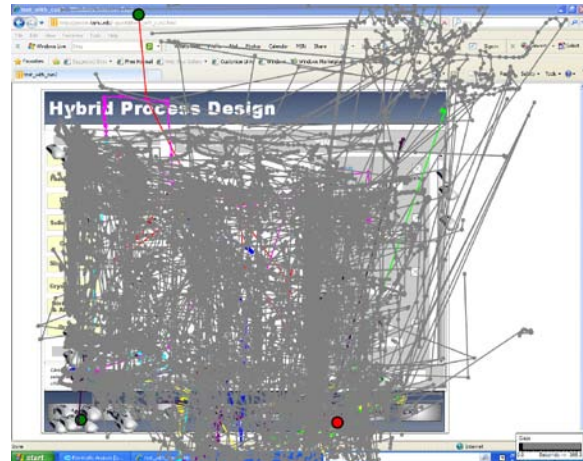


Fig. 4. Scanpath from experiment using Prototype I

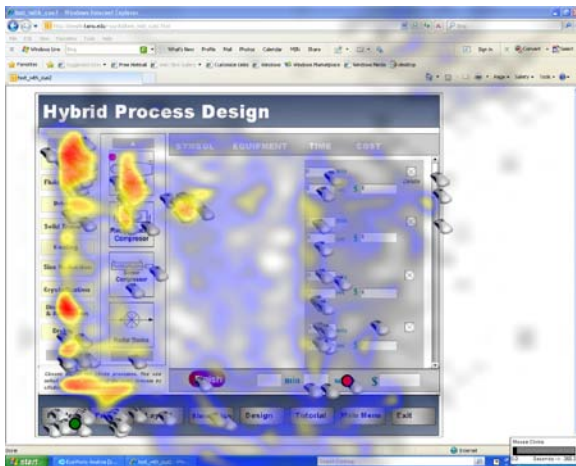


Fig. 5. Heat map from experiment using Prototype I

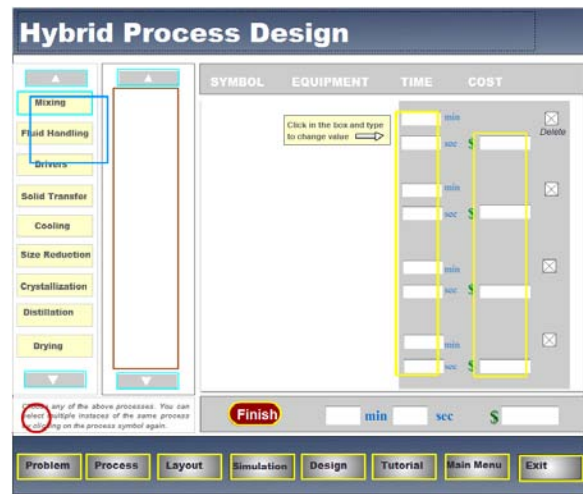


Fig. 6. Prototype II – Colored rectangles as visual cues

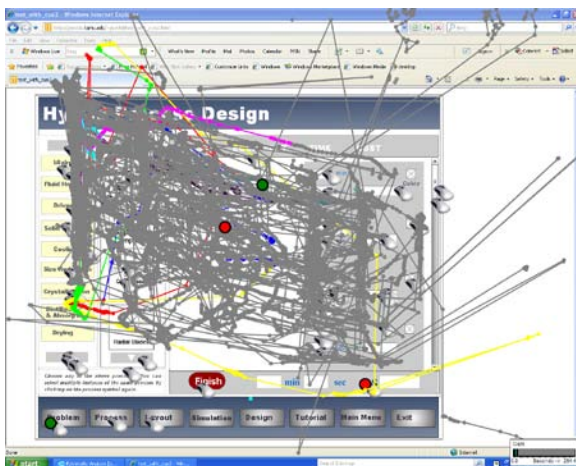


Fig. 7. Scanpath from experiment using Prototype II

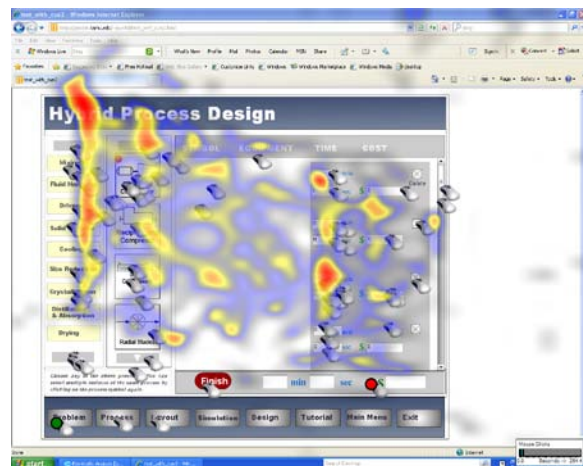


Fig. 8. Heat map from experiment using Prototype II

Prototype II

Prototype II has visual cues to help direct the viewer's gaze towards significant areas of interest. Hollow rectangles appear over the regions where the user is supposed to be looking next. The rectangles also appear around buttons that are important for the next step. For example, when the user clicks on the Mixing button, a rectangle appears over the region where mixing equipment is listed to aid with selection (Figure 6). Saturated colors were used for the rectangles since the color scheme used in the PSE consists of less saturated colors. Five participants were presented with this prototype. The results were slightly different from the experiment with Prototype I.

Figure 7 shows the scan path from one of the experiments on Prototype II. It is less cluttered as compared to the scanpaths from Prototype I as shown in Figure 4. The heat map in Figure 8 shows red areas over the Time and Cost regions.

Prototype III

This Prototype is a variation of Prototype II. Instead of colored rectangles, red dots serve as visual aids in this prototype. The appearance of the red dots follows the exact same pattern as that of the rectangles. Figure 9 shows an in-development screen shot of the application with red dots on the screen.

Figure 10 shows scan-paths from the experiment conducted on Prototype III. The scan-path is less cluttered and shorter in length as compared with the scan-paths obtained from the experiments using Prototype I and II. The heat map in Figure 11 shows equal distribution of gaze over the significant regions.

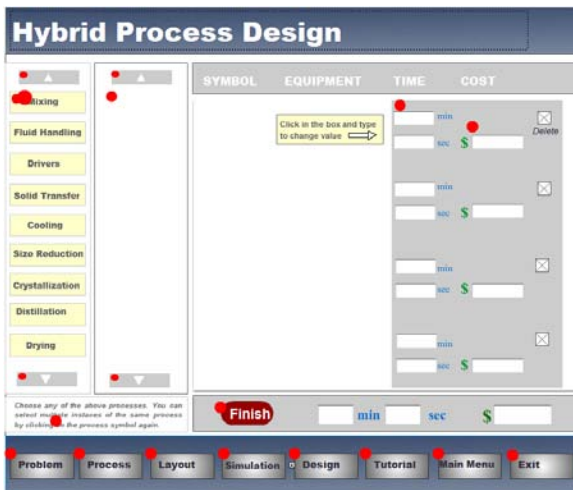


Fig. 9.: Prototype III – Red dots as visual cues

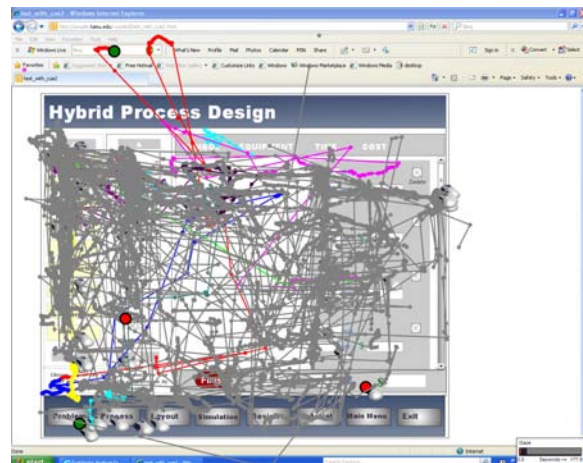


Fig.10.: Scanpath from experiment using Prototype III

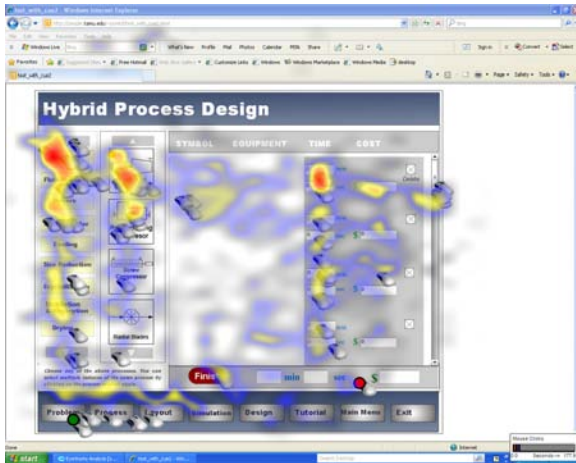


Fig. 11. Heat map from an experiment using Prototype III

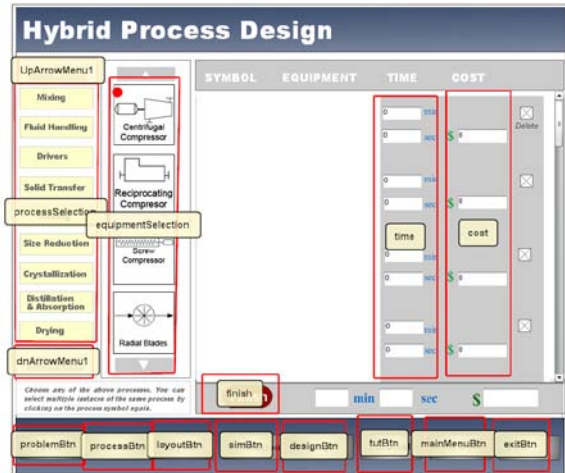


Fig. 12. Red rectangles represent regions for data analysis

Statistical Analysis

In order to generate statistics, it is necessary to identify the screen areas (regions) that need to be analyzed. Regions can be defined in the software used for analysis. The regions used for this study are shown in Figure 12. The areas marked with red rectangles are significant areas of interest. The software used for data analysis exports statistics according to the gaze hits and fixations received per region. The regions labeled Process Selection, Equipment Selection, Time, and Cost are the significant areas of interest to which users are supposed to be paying more attention. Table 1 shows gaze percentage in each region for Prototype I (No visual cues). The gaze percentages over the Cost and Time regions are significantly smaller in comparison to gaze percentages over the other regions. This further corroborates the observations made from heat maps in Figure 5, which shows that there are no red areas over these regions.

Table 1. Gaze percentage in each region for Prototype I

	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Cost	2.3	3.2	1.7	2	3.1
Equipment Selection	9.2	18.7	22.8	20.2	13.9
Process Selection	16.9	14	16	12.7	11.7
Time	6.7	8.6	8.2	7.4	7.7

Table 2 shows gaze percentage in each region for Prototype II (Rectangles). Gaze percentages in the Time and Cost regions have slightly improved for most of the participants. This means that the participants in Prototype II group (Rectangles) had their gaze directed towards the more significant areas more frequently than that in Prototype I (No cues) group.

Table 2. Gaze percentage in each region for Prototype II

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Cost	3.6	5	5.5	2.3	4
Equipment Selection	16.9	25.9	13	12.5	18.1
Process Selection	12.1	8.7	16.1	18.9	7
Time	6.3	12.4	12	16.7	6.8

Table 3 shows gaze percentage in each region for Prototype III (Red dots). Gaze percentages for the Time and Cost regions for the Prototype III group increased significantly over that of Prototype I and Prototype II. This shows that gaze was directed more towards the significant regions in Prototype III as compared to prototypes I and II.

Table 3. Gaze percentage in each region for Prototype III

	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Cost	1.2	5.8	6.3	7.6	3.8
Equipment Selection	16.2	16.6	21.9	11.7	16.9
Process Selection	10.2	10.5	14.5	10.9	11.8
Time	10.3	14.5	14.7	13.5	7.1

From Prototype I to Prototype III, gaze percentages for Time and Cost regions increase while gaze percentages for other regions decrease progressively. This trend can be clearly seen in Table 4, which compares averages of gaze percentages for the prototypes. The decrease in the gaze percentage for the Equipment Selection and Process Selection regions, coupled with the increase in gaze percentages for the Time and Cost regions means that participants looked at Time and Cost more frequently than Equipment Selection and Process Selection.

Table 4. Average gaze percentage in each region for all three prototypes

	Prototype I	Prototype II	Prototype III
Cost	2.46	4.08	4.94
Equipment Selection	16.96	17.28	16.66
Process Selection	14.26	12.56	11.58
Time	7.61	10.84	12.02

Figure 13 shows a graphical representation of the data presented in Table 4. The average gaze percentage for the Cost and Time regions increased from Prototype I (No cues) to Prototype III (Red dots). The visual cues in prototypes II and III were intended to draw the user's gaze towards previously neglected Cost and Time regions. Prototype III does a more efficient job of directing the user's gaze towards significant areas of interest as compared with Prototype II.

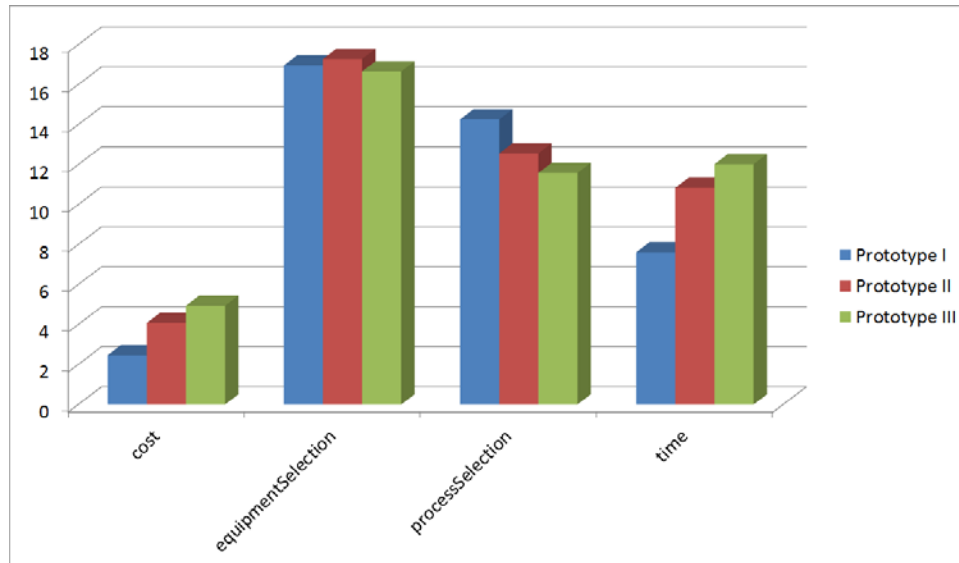


Fig. 13. Average gaze percentages for Cost, Equipment Selection, Process Selection and Time regions for all the prototypes

Statistical hypothesis testing (t-tests). To determine if the gaze percentages are significantly different among prototypes I, II and III, an independent samples t-test was performed on the data sets. Table 5 shows the probability values associated with t-tests for each region.

Table 5. Values for t-tests

	P1 (Prototypes I &II)	P2 (Prototypes I & III)
Cost	0.033906	0.06423
Equipment Selection	0.927688	0.920469
Process Selection	0.503498	0.064065
Time	0.15094	0.020768

P is the probability that the difference between two data sets is significant. The P-value for cost when comparing the Prototype 1 and II data sets, $P1(\text{cost}) = 0.033$. Because this value is less than 0.05 ($P1(\text{cost}) < 0.05$), we can conclude that there is a significant difference between the gaze percentages in the Cost region for Prototype I (No cues) and Prototype II (rectangles). The P-value for Cost when comparing the Prototype 1 and III data sets, $P2 = 0.064$, which is near to 0.05. The same is true for the p-value in all regions of Prototype III except for the Equipment Selection region. All the p-values for Prototype II (rectangles) are higher than 0.05 except for the Cost region. This implies that there is a significant difference between gaze patterns for Prototype I (No cues) and Prototype III (Red dots) as compared to the difference between gaze patterns between Prototype I (No cues) and Prototype II (Rectangles). Hence, the hypothesis that there is a difference between gaze patterns of Prototype I and III holds true, but it does not hold true for significant gaze pattern differences between Prototypes I and II.

Correlation coefficients. It is expected that Prototype III (red dots) does a better job of directing the user's gaze around the significant areas of interest as compared with Prototype II (rectangles). To determine this, gaze percentages for each region from four participants who used

Prototype II (Rectangles) were averaged. The coefficient of correlation was calculated between the average values and the gaze percentage values of a fifth participant from the same group. The coefficient of correlation in this particular case came out to be 0.54 ($r_1=0.54$). In the same way, the coefficient of correlation for Prototype III was calculated ($r_2=0.85$). The difference between r_1 and r_2 shows that there is greater linear correlation among the gaze percentages of participants in the Prototype III group than among the Prototype II group. Based on the earlier t-test results and the correlation results, it is deduced that gaze percentages in each region across all five Prototype III participants match closely with one another. A high degree of correlation among the gaze percentage of participants who used the same prototype indicates that most of the users followed the navigational path that their eyes were supposed to follow with the help of visual cues. The same cannot be said to be true for Prototype II (rectangles) which has a lower coefficient of correlation.

Tests of variance (F-tests). An F-test was performed on the data sets to determine if the variances of the two gaze percentage data sets were equal. Table 6 shows the F-test values.

Table 6. Values for f-tests

	F1 (Prototypes I & II)	F2 (Prototypes I & III)
Cost	0.248628	0.025214
Equipment Selection	0.994064	0.454009
Process Selection	0.140012	0.671259
Time	0.004559	0.013331

F1 and F2 are the probabilities that the variances are equal. F1 is the probability value calculated between gaze percentage for Prototype I and Prototype II, while F2 is the probability value calculated between gaze percentage for Prototype I and Prototype III. For the variances to be equal, the probability value should be equal to one. Table 6 shows that F2 is less than F1 except for the Process Selection and Time regions. The F1 and F2 values for all the regions are significantly less than one. This shows that the variance values for prototypes II and III differ from Prototype I.

Conclusion

Eye-tracking experiments were conducted with fifteen participants to evaluate the usability of the PSE. The following factors led to conclusions:

- Participants who tested Prototype III had smaller and less cluttered scan-paths as compared with that of Prototypes I and II.
- Prototype III had better distribution of gaze percentage over significant areas of interest as compared with Prototypes I & II.

These factors suggest that Prototype III (red dots) is the most navigable of the three prototypes. Prototype II, which has hollow rectangles as visual cues, performed better in terms of usability than Prototype I, which has no visual cues. The results recorded for the group that tested Prototype II were not consistent among all the five participants of the group, which means that not all of the participants of the group found it to be easily usable. The hypothesis “An e-learning

application which has visual cues to direct the user's gaze is easier to navigate, and is a more intuitive e-learning tool than the one without visual cues." was tested and determined to be true for both Prototype II (Rectangles) and Prototype III (Red dots).

Future directions include analyzing the number of mouse clicks per region to assess if there are differences in mouse click patterns among the three prototypes; the number of gaze observations and fixations registered by different subjects; the time to first fixation in each region; and the amount of time that participants took to finish the assigned task.

Acknowledgements

This material was supported by the National Science Foundation's Innovations in Engineering Education, Curriculum, and Infrastructure (IEECI) program (Award Number 0935211). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Bibliography

- [1] Hsieh, S. and *Deotale, P. "Design of Problem Solving Environment for Automated System Integration Education," 2011 ASEE Annual Conference, June 26-29, Vancouver, B.C., Canada.