Human Factors and Ergonomics in Manufacturing, Vol. 16 (1) 61–81 (2006)
DOI: 10.1002/hfm.20039

Applying the Proximity Compatibility and the Control-Display Compatibility Principles to Engineering Design Interfaces

Ling Rothrock, Kimberly Barron, Timothy W. Simpson, Mary Frecker, Chris Ligetti
The Harold and Inge Marcus Department of Industrial & Manufacturing Engineering, 210 Leonard Building, The Pennsylvania State University, University Park, PA 16802

Russell R. Barton
Department of Supply Chain and Information Systems, The Pennsylvania State University, University Park, PA 16802

ABSTRACT

The authors determine the utility of applying two display design principles toward the development of interfaces for engineering design. The first principle, called the *Proximity Compatibility Principle*, specifies that displays relevant to a common task or mental operation should be rendered close together in perceptual space. The second principle, called the *Control-Display Compatibility Principle*, stipulates that the spatial arrangement and manipulation of controls should be easily distinguishable. To examine the utility of both principles, the authors conducted an experiment comparing the ability of subjects to find effective designs using a separable versus a configural interface in a multi-objective engineering design task. Results suggest that the proximity compatibility principle is an effective indicator of task performance. Moreover, the control-display compatibility principle can be used as an indicator of performance efficiency. © 2006 Wiley Periodicals, Inc.

1. INTRODUCTION

The state-of-the-art in visualization to support engineering design decisions is still considered to be in its infancy (Messac & Chen, 2000). Jones (1994) argues that appropriate representations (i.e., visualization strategies) are needed to better understand the models, algorithms, data, and design candidates obtained during the design process. A recent study conducted by the National Research Council (NRC) highlighted three requirements for an effective user interface for engineering design: It must be integrative, visual, and *fast*, in order to enable real-time response to user input (National Research Council, 1998).

A user interface, which Norman (1991) calls an “artifact,” mediates between the user and the world both in terms of execution (between actions and the resulting changes to the world state) and perception (between changes in the world and our interpretation

Correspondence to: Ling Rothrock, Department of Industrial & Manufacturing Engineering, 210 Leonhard Building, The Pennsylvania State University, University Park, PA 16802. E-mail: lrothroc@psu.edu

of it). Norman's distinction is important in the context of engineering design because designers must bridge the gulf that separates their goals and intentions from the world. In the context of computer-based engineering design, the world is encapsulated in a computer which analyzes and renders the designs.

A user interface must accomplish two objectives: (a) make available information required by designers to assess the state of the design with reference to their goals; and (b) enable designers to execute actions needed to accomplish those goals (Hutchins, Hollan, & Norman, 1986). With the advent of computer-based engineering design, it is critical that design interface concepts complement technological advances to optimize the potential of *metamodels* toward generating effective and efficient solutions. By metamodels, we mean simple mathematical approximations to the input/output functions calculated by the designer's analyses and simulation models (Barton, 1998; Simpson, Peplinski, Koch, & Allen, 2001). An interface should extract the critical features of the design space to enable designers to attain their goals (Vicente & Rasmussen, 1992). Moreover, the interface's control-display relationship should be consistent with human perceptual and cognitive abilities (Karwowski, 2000; Tullis, 1988; Wickens, 1992) so that effective control of the design process can be achieved.

1.1. The Gulf of Evaluation

Interface design research has focused on the design of visual representations that enable the user to meet the cognitive load of the task (Bennett, Toms, & Woods, 1993; Carswell & Wickens, 1990; Pomerantz, 1986; Sanderson, Flach, Buttigieg, & Casey, 1989; Wickens & Carswell, 1995). Laying the groundwork for the field, Pomerantz (1986) proposed three types of relationships between visual stimuli: "separable," "integral," and "configural." A separable relationship is characterized by no interactions among stimulus dimensions (e.g., the perception of color does not affect the perception of length). An integral relationship, on the other hand, is defined by a strong interaction among dimensions so that individual dimensions are no longer perceptible (e.g., the dimensions of hue and brightness combine to form color). A configural relationship refers to an intermediate level of interaction between perceptual dimensions where each dimension maintains its unique perceptual identity while new, emergent properties are created because of the interaction between them (e.g., a pair of bars on a vertical bar graph reveal individual values while the difference in height between the bars create the perception of a slope—the emergent feature).

The importance of interface design research lies in the principle that defines relationships between task demands and the graphical form of a display (i.e., separable, integral, or configural). Conceived by Wickens and Carswell (1995), the principle, called the *Proximity Compatibility Principle* (PCP), specifies that displays relevant to a common task or mental operation should be rendered close together in perceptual space (Barnett & Wickens, 1988; Wickens & Carswell, 1995).

The PCP depends critically on two dimensions of similarity: display proximity and mental proximity. Display proximity refers to the perceptual similarity between information sources in a display and is defined along several dimensions: spatial proximity (physical distance); chromatic distance (same or different colors); code homogeneity (coding variable using the same or different properties such as length or color); and geometric form (integral or configural versus separable displays). For instance, if individual variables are mapped onto a geometric object, the display is high in display proximity. However, when each variable has its own separate representation (e.g., a digital readout), the display is low in proximity.

Mental proximity refers to the extent to which information from the various sources in a display must be considered together to accomplish a task. There are three categories of mental proximity: integrative processing (highest proximity), nonintegrative processing (intermediate proximity), and independent processing (lowest proximity). In integrative processing, information from multiple sources must be explicitly combined (e.g., computational processing involving numerical operations). In nonintegrative processing, features of similarity are categorized (e.g., statistical similarity in terms of the extent of covariation). Independent processing involves no interaction between information sources.

The PCP maintains that efficient interaction occurs when display proximity matches task proximity. Performance on integrated tasks (high mental proximity) is predicted to be facilitated by displays that have high perceptual proximity (an integral or configural display). Similarly, performance on focused tasks (low mental proximity) is predicted to be facilitated by displays that have low perceptual proximity (e.g., a digital readout; Bennett, Nagy, & Flach, 1997).

1.2. The Gulf of Execution

While PCP specifies display constraints so that efficient interaction can occur, it does not specify the form of the interface to enable execution. Nevertheless, much of the display control literature (Bullinger, Kern, & Braun, 1997; Wickens, 1992) can be distilled into some general design principles. Called the *principles of control-display compatibility* (Wickens, 1992), these stipulate that:

- The spatial arrangement of controls should allow users to easily tell which control is used.
- The indicator of a display should move in the same direction as its control.
- The layout of the operational method of controls should be consistent with expectations of the user population.
- The direction in which a part moves on the display should be consistent with user expectations.

The *control-display compatibility principle* (CDCP) serves as a guide to designing controls for a computer-based design interface. Therefore, as the input display represents a mapping from multiple mathematical dimensions (i.e., design variables) to graphical dimensions, the representation of changes in design variable values must be consistent with user expectations.

1.3. Displays in Design Tasks

Existing user interface research in manufacturing design contexts has focused primarily on the development of visual representations. Petre and Green (1992) interviewed practicing engineers and found that, for computer-aided design drawings, graphics are better used for overviews, zooming should be used for extracting detail, and high spatial proximity and connectedness between display elements are needed to provide structure. Beard and Walker (1990) showed that using a map window significantly improved navigation performance in a large two-dimensional design space. Gerace and Gallimore (2001) found that, for an assembly sequence planning task, integral displays enabled faster and more accurate diagnostic performance than separable displays. Hirschi and Frey (2002) showed that the time required to solve a problem requiring integrative processing using a separable display increases geometrically as the number of integrative parameters increases.

Our research premise is that effective interfaces for computer-based engineering design must enable the designer to effectively bridge the gulfs of evaluation and execution (Norman, 1991). We propose that the gulf of evaluation is effectively crossed using an interface in which the display proximity corresponds to the mental proximity demanded by the task. In the case of engineering design, a high mental proximity should require high display proximity. To bridge the gulf of execution, we submit that not only is proximity compatibility required but control-display compatibility is also necessary.

2. METHODS

2.1. Experimental Task Overview

The task used for our research is the design of a cantilever I-beam example adapted from Hafta and Gürdal (1992). The user must search for an optimal design to minimize the cross-sectional area (A) subject to an upper limit on the imposed maximum bending stress (σ) of an I-beam (see Figure 1) by adjusting the height (h) and width (w) of the I-beam cross-section.

Mathematically, the user is asked to solve the following constrained optimization problem:

$$\begin{array}{ll} \text{Minimize} & A \\ \text{Subject to:} & \sigma \leq \text{upper bound} \\ & 0.2 \leq h \leq 10.0 \\ & 0.1 \leq w \leq 10.0 \end{array}$$

The measures of A and σ are normalized so that they are of the same relative magnitude.

2.1.1. I-beam design interfaces. The I-beam design interfaces differ from each other in two ways: (a) the method of entering input variables h and w, and (b) in the visual

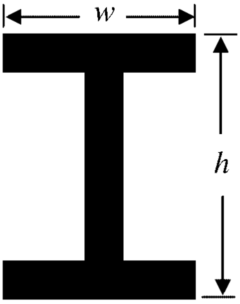


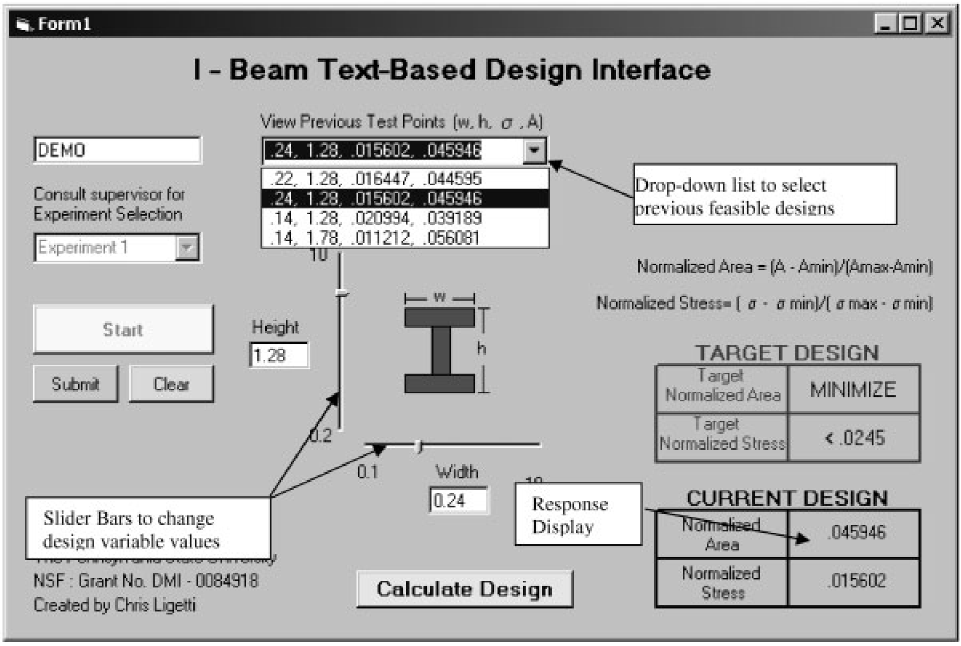
Figure 1 Cross-section of I-beam.

representation of response parameters A and σ. Input methods include slider bars for each input variable as well as a “fieldbox” with which the user can simultaneously vary both input variables. The response of the designs were either presented graphically on a 2D plot of A versus σ (response space plot), or numerically in a drop-down list. Each interface is discussed in detail in the following sections.

2.1.2. Separable interface. The separable interface (SI) is a modified version of the text-based design interface from Ligetti, Simpson, Frecker, Barton, and Stump, (2003). To ensure that all combinations of h and w were selectable, we used slider bars as the method to enter values for the input variables. To vary the input variable values of the I-beam, users manipulate a mouse to move the corresponding slider bar (horizontal to vary w, and vertical to vary h) as shown in Figure 2.

Designs are evaluated when the user presses the Calculate Design button on the bottom of the design interface. The constraint on σ is shown in the Target Design table on the right of the interface. When the Calculate Design button is pressed, the response corresponding to the current h and w is shown in the Current Design table. Designs that satisfy the constraint on σ are stored in a drop-down list of previously evaluated I-beam designs, where the w, h, σ, and A values are listed for each design. To remove all previous designs from the drop-down list, the user presses the Clear button. To submit the current design as the optimal for the I-beam design, the user presses the Submit button.

The interface is labeled as separable because only one arguably weak interaction exists between visual stimulus dimensions (w, h, σ, and A)—that of the drop-down list of previous designs. Otherwise, the information is coded as purely numerical data. Recall that interaction, in the current context, pertains to relationships between visual stimuli.



I - Beam Text-Based Design Interface

View Previous Test Points (w, h, σ, A)

.24, 1.28, .015602, .045946

.22, 1.28, .016447, .044595
.24, 1.28, .015602, .045946
.14, 1.28, .020994, .039189
.14, 1.78, .011212, .056081

Drop-down list to select previous feasible designs

DEMO

Consult supervisor for Experiment Selection

Experiment 1

Height 1.28

Width 0.24

Normalized Area = (A - Amin)/(Amax-Amin)

Normalized Stress= (σ - σ min)/(σ max - σ min)

TARGET DESIGN

Target Normalized Area	MINIMIZE
Target Normalized Stress	< .0245

Slider Bars to change design variable values

Response Display

CURRENT DESIGN

Normalized Area	.045946
Normalized Stress	.015602

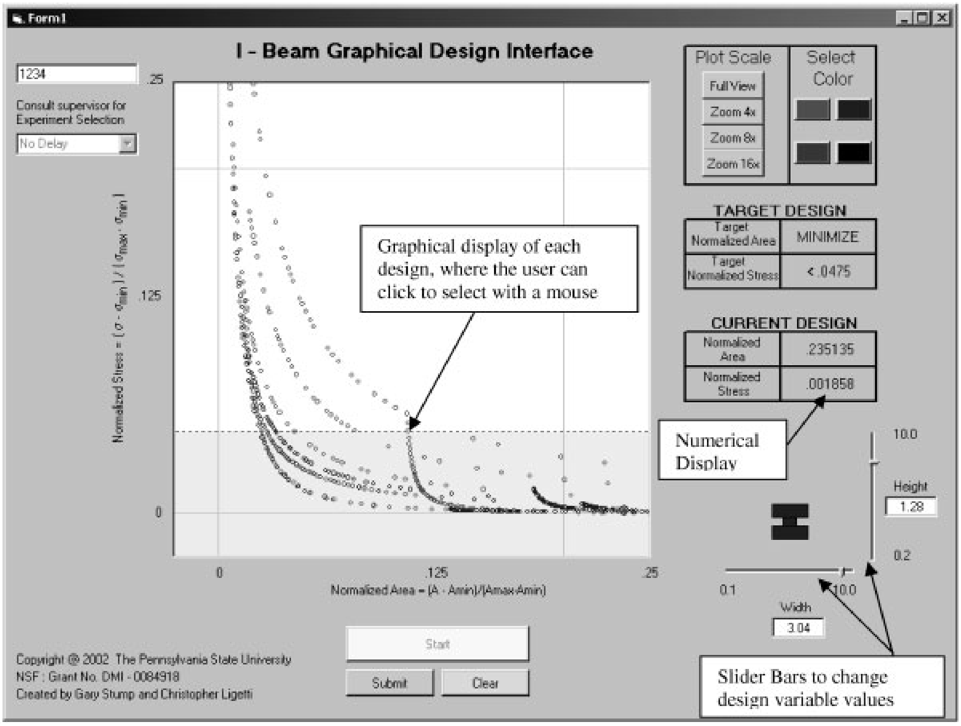
NSF : Grant No. DMI - 0084918
Created by Chris Ligetti

Figure 2 Separable interface layout and functionality.

In terms of the control-display compatibility, the separable interface fails to meet the principles stipulated by Wickens (1992) in two areas. First, the indicator of the display, shown as the I-beam in Figure 2, does not reveal the corresponding response of selected *w* and *h* values. Second, I-beam design changes in response to slider bar manipulations are not perceptually available to the user.

2.1.3. Configural interface-A. The design variable input method used in the configural interface-A (CI-A) is identical to that of the SI (as shown in Figure 3). The primary difference between the CI-A and the SI is that the response of the designs is presented graphically. The visual representation of the responses is shown in the form of a normalized A versus normalized σ plot. By manipulating input variables via slider bars, the user is provided with a 2D graphical representation of a corresponding I-beam design. When a design point is selected, its normalized A and σ values are displayed in the Current Design table to the right of the interface. CI-A allows the user to change the scale of the plot of A versus σ by selecting any of the four Plot Scale buttons that appear in the upper right portion of Figure 3. To remove all design points from the plot, the user presses the Clear button. To submit the current design as the optimal for the I-beam design, the user presses the Submit button.

The interface is labeled as configural because each dimension (w, h, σ, and A) maintains its unique identity while creating emergent properties based on interactions with other dimensions. Aside from the four input and response dimensions, five additional dimensions emerge from the two-dimensional (2D) plot:



1234

No Delay

TARGET DESIGN

Target Normalized Area	MINIMIZE
Target Normalized Stress	< .0475

CURRENT DESIGN

Normalized Area	.235135
Normalized Stress	.001858

Height 1.28

Width 3.04

Graphical display of each design, where the user can click to select with a mouse

Numerical Display

Slider Bars to change design variable values

Copyright @ 2002 The Pennsylvania State University
NSF : Grant No. DMI - 0084918
Created by Gary Stump and Christopher Ligetti

Figure 3 Configurable interface-A layout and functionality.

1. $\sigma \leq$ upper bound, the bending stress constraint.
2. $\Delta\sigma$, the difference between the bending stress of the previous response and the current.
3. ΔA, the difference between the cross-sectional area of the previous response and the current.
4. $d((A_t, \sigma_t), (0, 0))$, the distance between the current response and the origin.
5. $d((A_t, \sigma_t), (A_{t-1}, \sigma_{t-1}))$, the distance between the current response and the previous response.

We focus on these five additional dimensions because decision-making behavior is often guided not by the current state occupied by the system but by the visual perception of the system's state transitions (Gibson, 1986).

With respect to control-display compatibility, the CI-A does much better than the SI in meeting Wickens' principles. The display indicates both the state of the input variables as well as the graphical representation, in 2D space, of the corresponding response. Moreover, manipulation of one of the input variables causes a change (either Δh or Δw) that is reflected in the design parameters so that $\Delta h \rightarrow (\Delta\sigma, \Delta A)$ or $\Delta w \rightarrow (\Delta\sigma, \Delta A)$. Therefore, as users become more practiced on the task, the functional relationship between input and response variables should become more familiar.

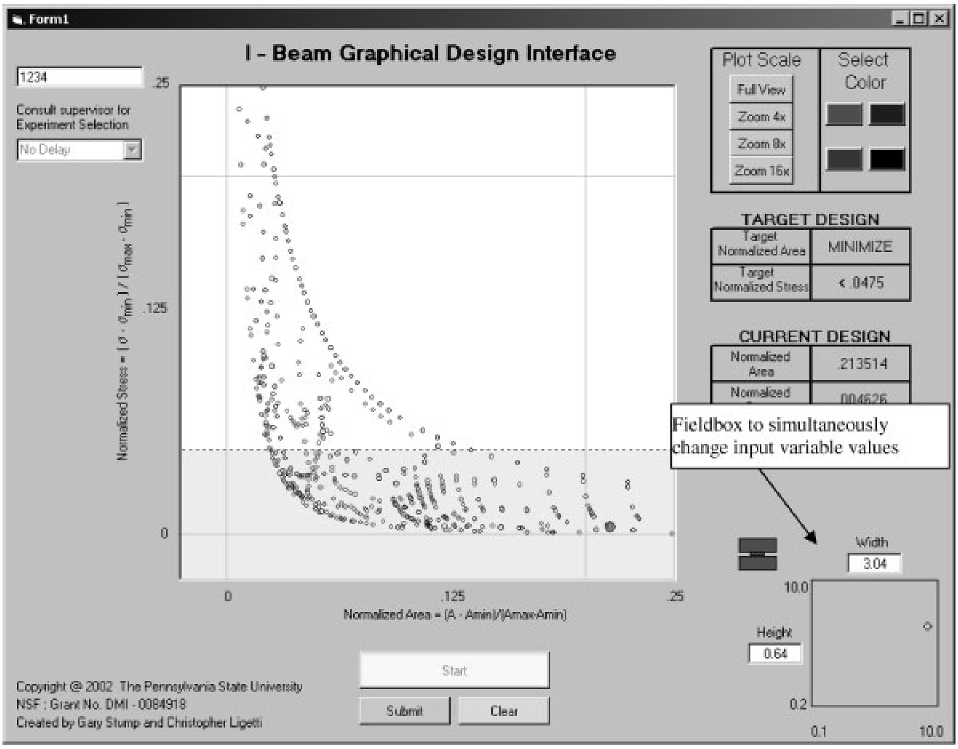
2.1.4. Configural interface-B.

The difference between CI-A and CI-B lies in the method for entering the input variables. Configural interface-B allows for simultaneous input variation using what we call a "fieldbox." The fieldbox is simply a box with w and h on the horizontal and vertical axes, respectively, and a cursor inside the box, as shown in Figure 4. The user performs a click-and-drag operation to change the cursor position, and the corresponding w and h values, anywhere within the boundaries of the box. In terms of control-display compatibility, the CI-B differs from CI-A in that the functional relationship between the changes in *both* input variables and corresponding change in response space is perceptually available. We represent this relationship as $(\Delta h, \Delta w) \rightarrow (\Delta\sigma, \Delta A)$.

2.1.5. Mappings for domain properties and constraints.

The description of each interface revealed several differences with respect to a user's ability to bridge the gulfs of evaluation and execution. We provide a more concise comparison (see Table 1) by listing differences between the three interfaces in terms of input and response variables that are directly perceivable versus those that must be derived by the designer. Note that distance and constraint information in the last three rows of Table 1 are not perceptually available to users of the SI. Furthermore, the process of deriving the information will require additional cognitive resources. We use the visual distinctions between the interfaces to generate our hypotheses for the applicability of the PCP toward engineering design.

Complementing the comparisons in design representation, Table 2 shows differences among the three interfaces in how the functional relationship between input and response parameters are displayed and manipulated. The distinctions between the interfaces are used to formulate our hypotheses for the applicability of control-display compatibility toward engineering design. It is possible for the user in the SI condition to make the $\Delta h \rightarrow (\Delta\sigma, \Delta A)$ or $\Delta w \rightarrow (\Delta\sigma, \Delta A)$ relationships perceptually available by repeatedly clicking the Calculate Design button. However, it is doubtful that the $(\Delta h, \Delta w) \rightarrow$



1234

Consult supervisor for Experiment Selection

No Delay

TARGET DESIGN

Target Normalized Area	MINIMIZE
Target Normalized Stress	< .0475

CURRENT DESIGN

Normalized Area	.213514
Normalized	.004626

Width 3.04

Height 0.64

Copyright @ 2002 The Pennsylvania State University
NSF : Grant No. DMI - 0084918
Created by Gary Stump and Christopher Ligetti

Figure 4 Configural interface-B layout and functionality.

$(\Delta\sigma, \Delta A)$ relationship can be established by the user in the SI. The key difference between CI-A and CI-B lies in the perceptual availability of sequentially manipulating individual input variables using slider bars versus simultaneously manipulating both input variables using the fieldbox.

TABLE 1. Visual Display Mappings for Design Parameters and Constraints

	Separable interface		Configural interfaces	
Parameters	Perceivable	Derivable	Perceivable	Derivable
h	X		X	
w	X		X	
σ	X		X	
A	X		X	
$\Delta\sigma$	X		X	
ΔA	X		X	
$d((A_t, \sigma_t), (0,0))$		X	X	
$d((A_t, \sigma_t), (A_{t-1}, \sigma_{t-1}))$		X	X	
$\sigma \leq$ upper bound		X	X	

TABLE 2. Visual Display Mappings for Input Parameters and Controls

	Separable interface		Configural-A interface		Configural-B interface	
Parameters	Perceivable	Derivable	Perceivable	Derivable	Perceivable	Derivable
h	X		X		X	
w	X		X		X	
$\Delta h \rightarrow (\Delta\sigma, \Delta A)$		X	X			X
$\Delta w \rightarrow (\Delta\sigma, \Delta A)$		X	X			X
$(\Delta h, \Delta w) \rightarrow (\Delta\sigma, \Delta A)$				X	X	

2.2. Research Hypotheses

We decompose our research premise into three major hypotheses to investigate the applicability of the proximity compatibility and the control-display compatibility principles in computer-based engineering design. To investigate the robustness of the interfaces, we also varied the level of task load in terms of the time delay between the manipulations of the input variables and the presentation of the corresponding response variables. Detailed analyses of the impact of time delay on user performance are discussed in (Barron et al., 2005).

Proximity Compatibility Principle Hypothesis: An interface with high display proximity is better matched with the I-beam design task than an interface with low display proximity in normal and elevated task load conditions.

Corollary: Proximity compatibility is persistent under varying conditions of task load. Subjects trained under a normal task load should therefore be affected similarly, according to the Interface Type, by elevated task loads.

Control-Display Compatibility Principle Hypothesis: A display with high control-display compatibility enables more efficient performance than a display with low compatibility in normal and elevated task load conditions.

Corollary: Control-display compatibility is persistent under varying conditions of task load.

Workload Hypothesis: There exists an inverse relationship between user-perceived workload in the I-beam design task and the PCP and CDCP. We propose that as the correspondence between display proximity and mental proximity increases, the user-perceived workload should decrease. Similarly, as the control-display compatibility increases, user-perceived workload should also decrease.

2.3. Subjects

Sixty Pennsylvania State University students (43 men and 17 women) participated in the experiment. Participants were 18 to 34 years of age ($M = 24.52$) and reported weekly computer usage of 5 to 70 hours ($M = 28.67$). The experiment took approximately 1 hour to complete, and subjects were compensated $20 for their time and effort.

2.4. Experiment Design

A mixed-factors design with one within-subject (trial) and two between-subjects factors was employed with 10 replications. One between-subjects factor consisted of interface type (SI, CI-A, or CI-B) and the other included task load (no time delay for normal load or 1.5 seconds time delay for elevated load). The "time delay" represents the time elapsed from the user when the user submits a design to when the performance is displayed. The 1.5 s delay was used by Goodman and Spence (1978) to provide a noticeable effect in the behavior of the interface.

The dependent measures in the experiment include the percent error between each submitted design and the known optimal design; the completion time for each design task; and the number of designs explored prior to submission. We also examined the perceived user workload using the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). The NASA-TLX is a multidimensional rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales: Mental Demands (level of mental and perceptual activity required), Physical Demands (level of physical activity required), Temporal Demands (amount time pressure felt), Own Performance (level of perceived successfulness), Effort (level of work required), and Frustration (amount of discouragement, irritability, and stress felt). It is a widely used subjective workload measure that provides the users with a direct method of providing their opinions, has high face validity, and has been shown to be sensitive to a variety of task commands (Wierwille & Eggemeier, 1993).

2.5. Pilot Study

Prior to recruiting subjects, we conducted a pilot study to assess the sensitivity of our design. In particular, we sought to find the number of subjects needed to achieve a power of 0.80 based on a two-tailed α of 0.05. Twelve subjects participated in the pilot study in which a multifactor design was employed where interface type and time delay were between-subject variables. To ascertain the number of training trials required for skill acquisition, each subject ran 10 trials. The results of the pilot study in terms of the percent error of the submitted design and the time required to submit the design are shown in Figure 5. In both figures, a quadratic fit was applied to indicate the learning curve.

To determine the number of subjects required for sufficient power, and to mitigate Type II errors, we apply the power analysis for analysis for main effects and interactions introduced by Cohen (1988). Based on the learning curve shown in Figure 5, we selected trials 6 to 8 for our power calculations. Therefore, in the case of the percent error for $\alpha = 0.05$ on the time delay factor and 0.80 power, the number of subjects (n) needed is approximately 48 for an effect size (f) of 0.29. For the percent error on the interface type factor at $\alpha = 0.05$ and $f = 0.31$, $n \approx 48$. In the case of completion time on the time delay factor at $\alpha = 0.05$ and $f = 0.34$, $n \approx 48$. For completion time on the interface type factor at $\alpha = 0.05$ and $f = 0.35$, $n \approx 40$.

The results of the pilot study suggest that at least 48 subjects are needed for the experiment. Moreover, a visual inspection of the learning curve suggests that eight trials are sufficient for reaching task proficiency.

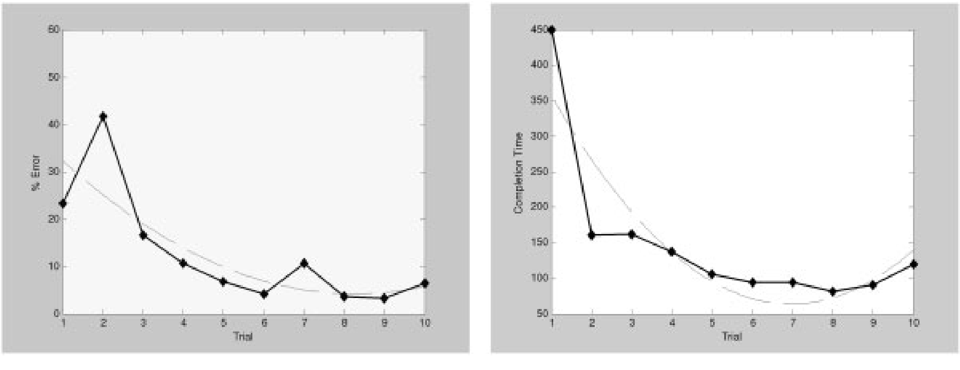


Figure 5 Pilot test results for 10 trials measured in percent error and completion time.

2.6. Procedure

Each participant was randomly assigned to one of the interface types and time delays. Participants were given a packet of instructions that included a description of the experiment and a one-page overview of the I-beam design problem. They also completed a questionnaire to provide demographic data at this time.

Prior to starting the first trial, the experiment facilitator instructed the participant to find the best design in the shortest time possible. To begin each trial, a participant pressed the Start button (see Figures 2–4). When the participant was satisfied with the final design, the Submit button was pressed to indicate his or her selection and to proceed to the next trial. Completion time represents the time elapsed between pressing the Start and Submit buttons.

Each participant ran 12 trials of the assigned interface type and time delay condition. The first eight trials were used to train the subject to perform at a criterion level based on results from the pilot study. Data from the 9th and 10th trials were recorded to test task performance. Finally, the 11th and 12th trials were used to examine the transfer effects of moving to a reverse-delay condition (i.e., subjects with the 1.5 s delay were given no delay and vice versa).

After each trial, the NASA-TLX was administered via computer. Following the completion of the final trial, participants completed a postexperiment questionnaire to indicate perceived performance in the experiment as well as any problems that were encountered during the experiment. A graduate student facilitator supervised each experiment, administered the questionnaires and NASA-TLX, and answered questions.

3. RESULTS

We found that the natural logarithm of the percent error tended to produce more normally distributed variation; hence, the natural logarithm of the percent error, ln(error), was used for analysis. Similarly, the natural logarithm of task completion time, ln(completion time), was also used.

3.1. Proximity Compatibility Principle Hypothesis Testing

Results that address the PCP Hypothesis are shown in terms of accuracy (% error from the known optimal design) in Figure 6 and in terms of latency (completion time for generating design) in Figure 7.

To determine accuracy and latency, we ran Analyses of Variance (ANOVAs) using time delay (0.0 or 1.5) and interface type (SI, CI-A, or CI-B) as between-subjects factors and Trial (9 and 10) as the repeated measures variable. In terms of accuracy (see Figure 6), the main effect of time delay was significant, $F(1,54) = 17.29$, $p < .000$, but interface type was not significant, $F(2,54) = 0.52$, $p < .771$. Moreover, the interaction effect of interface × delay was not significant, $F(2,54) = 1.53$, $p < .226$.

From the perspective of latency (Figure 7), both the main effects of time delay, $F(1,54) = 5.22$, $p < .026$, and interface type, $F(2,54) = 4.39$, $p < .017$, were significant. The interface × time delay was not significant at $F(2,54) = 2.16$, $p < .125$. To further explore the significance among the different interfaces, we conducted pairwise comparisons among all three types of interfaces using the Tukey HSD procedure. The results of this analysis

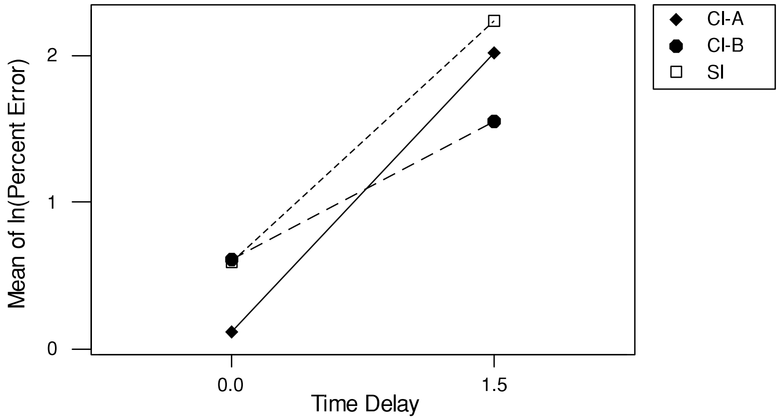


Figure 6 Mean transformed percent error scores for normal and elevated task load conditions.

indicate that the group using CI-B improved completion time significantly more than either the CI-A or SI groups.

To examine the corollary to the PCP hypothesis, we analyzed subject performance under reverse-delay conditions. We again used accuracy and latency as the dependent variables and ran ANOVAs using interface type (SI, CI-A, or CI-B) as between-subjects factors and trial (9–12) as the within-subject factor. First, data from the subject group trained and tested under normal task load conditions was analyzed in terms of accuracy.

In terms of accuracy (see Figure 8), the main effect of trial was significant, $F(3,81) = 22.58$, $p < .000$, but interface type was not significant, $F(2,27) = .630$, $p < .540$. Moreover, the interaction effect of trial × interface was not significant, $F(6,81) = .583$, $p < .743$. These results suggest that elevated task load alone impacted subject performance in terms of increased errors.

From the perspective of latency (see Figure 9), both main effects of trial, $F(3,81) = 17.14$, $p < .000$, and interface type, $F(2,27) = 5.80$, $p < .008$, were significant. How-

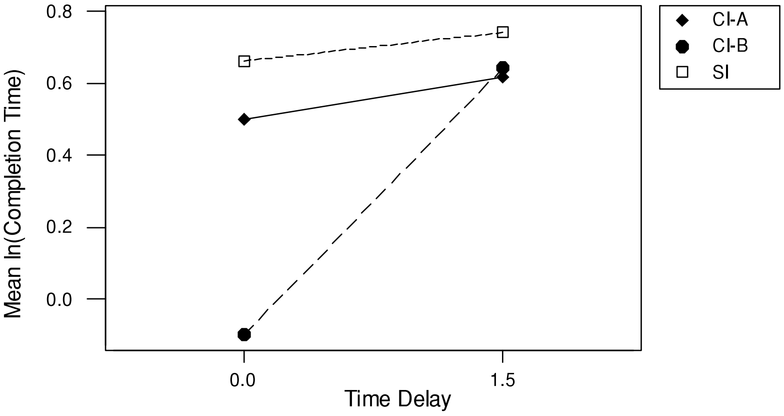


Figure 7 Mean transformed times for normal and elevated task load conditions.

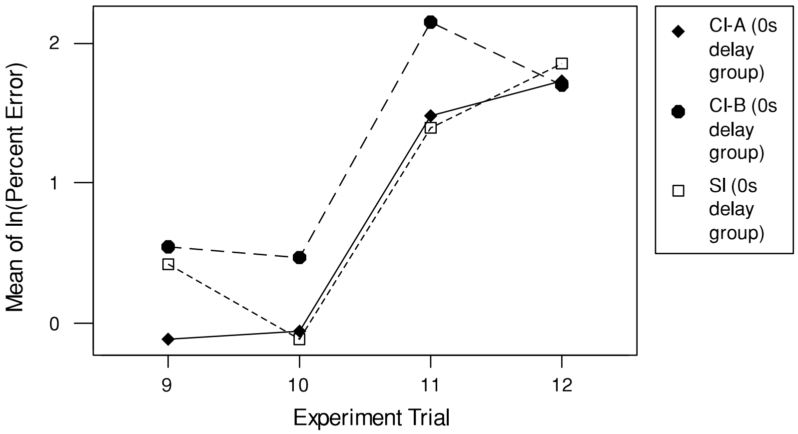


Figure 8 Mean transformed percent error values for test (Trials 9–10) and transfer (Trials 11–12) phases of the experiment. Each subject group was trained and tested on an assigned interface with 0 s delay and then transferred to the same interface with 1.5 s delay.

ever, trial × interface was not significant, $F(6,81) = 1.57, p < .167$. Therefore, evidence suggests that while both elevated task load and CI-A and SI conditions contributed toward increased completion times, performance on different interfaces did not depend on the task load condition.

We also analyzed subject performance under reverse-delay conditions in the opposite order. Data from the subject group trained and tested under the elevated task load condition was collected for the normal task load condition.

In terms of accuracy (see Figure 10), the main effect of trial was significant, $F(3,81) = 3.351$, $p < .023$, but interface type was not significant, $F(2,27) = 2.243$, $p < .126$. In addition, the interaction effect of trial × interface was significant, $F(6,81) = 2.673$,

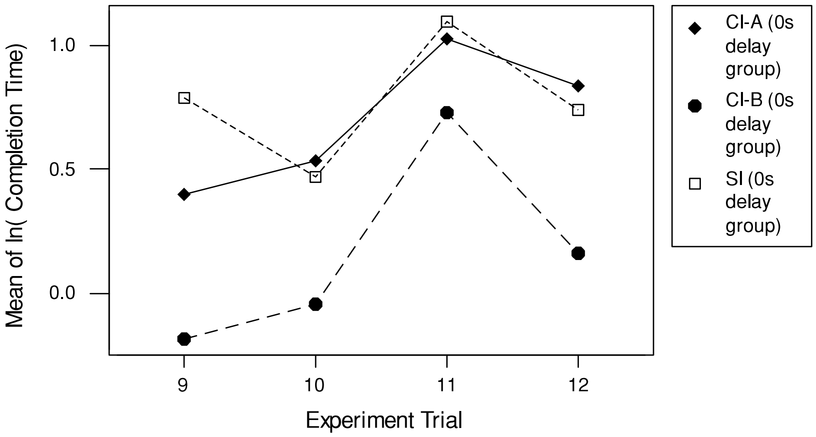


Figure 9 Mean transformed completion times for test and transfer phases of the experiment. Each subject group was trained and tested on an assigned interface with 0 s delay and then transferred to the same interface with 1.5 s delay.

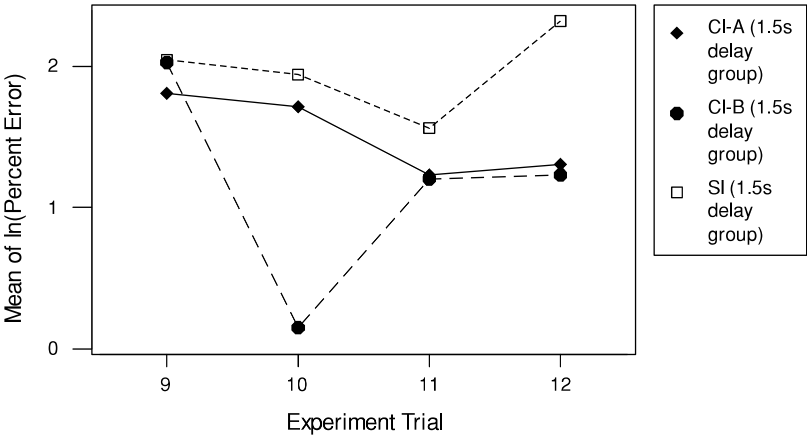


Figure 10 Mean transformed percent error values for test and transfer phases of the experiment. Each subject group was trained and tested on an assigned interface with 1.5 s delay and then transferred to the same interface with 0 s delay.

$p < .020$. A visual inspection of the significant interaction effect suggests that introducing normal task load after training on the elevated task load condition causes the highest performance disruption between Trials 10 and 11. To investigate the interaction significance in terms of trials, we used a within-subject contrast, or comparison, of trial means in the following form:

$$\Psi_T = c_{T9}\mu_{T9} + c_{T10}\mu_{T10} + c_{T11}\mu_{T11} + c_{T12}\mu_{T12} \tag{1}$$

where μ_T is the mean for trial T. In particular, we found that for the contrast, $\Psi_T = 0\mu_{T9} + 1\mu_{T10} - 1\mu_{T11} + 0\mu_{T12}$, to be significant, $F(2,27) = 4.02, p < .030$. More importantly, this result highlights the finding that a shift from the elevated task load to normal task load increases errors in the CI-B condition.

In terms of latency (see Figure 11), the main effect of trial, $F(3,81) = 3.158, p < .029$, was significant but interface type, $F(2,27) = .191, p < .827$, was not. The trial × interface interaction was not significant, $F(6,81) = .821, p < .557$. These results suggest that elevated task load alone affected subject performance in terms of longer completion times.

3.2. Control-Display Compatibility Principle Hypothesis Testing

To test the CDCP Hypothesis, we focused our analysis on the quantity of potential designs explored in each of the interfaces. A design candidate is generated for the SI condition when a subject clicks the Calculate Design button (see Figure 2). A potential design is evaluated for the CI conditions when a subject clicks on a response point in the normalized plot of A versus σ (see Figures 3 and 4).

To analyze design execution effectiveness, we used the quantity of designs explored as the dependent variables and repeated the analyses conducted for testing the PCP Hypothesis. First, we ran an ANOVA using time delay (0.0 or 1.5) and interface type (SI, CI-A, or CI-B) as between-subjects factors and trial (9 and 10) as the within-subject factor. A

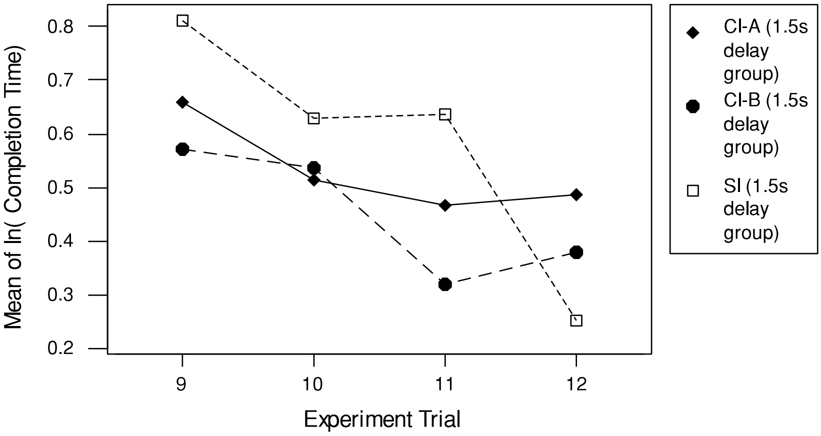


Figure 11 Mean transformed completion times for test and transfer phases of the experiment. Each subject group was trained and tested on an assigned interface with 1.5 s delay and then transferred to the same interface with 0 s delay.

plot of means is shown in Figure 12. The main effect of interface type was significant, $F(2,54) = 14.86, p < .000$, but time delay, $F(1,54) = .848, p < .361$, and trial, $F(1,54) = 1.504$, $p < .225$, were not significant. None of the interaction effects was significant. To follow up on the significance among interfaces, we conducted pairwise comparisons among all three types of interfaces using the Tukey HSD procedure. The results of this analysis indicate that the group using SI explored a significantly larger number of designs.

To investigate the corollary to the CDCP Hypothesis, subject performance under reverse-delay conditions was analyzed. We again used quantity of designs explored as the dependent variables and ran ANOVAs using interface type (SI, CI-A, or CI-B) as between-subjects factors and trial (9–12) as the within-subject factor. First, data from the subject group trained and tested under normal task load conditions and then transferred to the elevated task load condition was analyzed (see Figure 13). Only the main effect of inter-

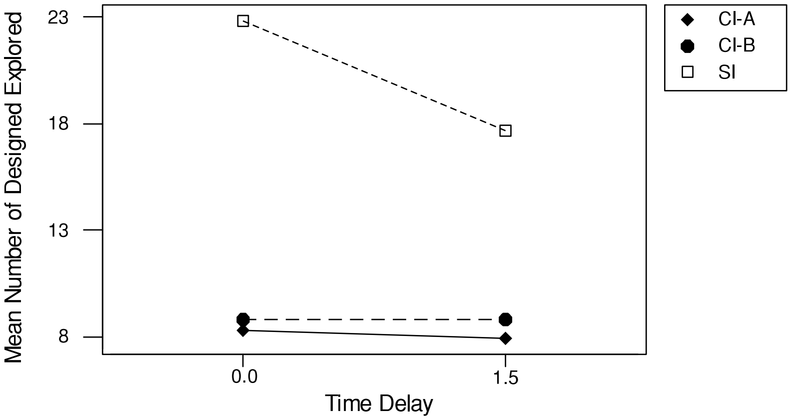


Figure 12 Mean number of designs explored for normal and elevated task load conditions.

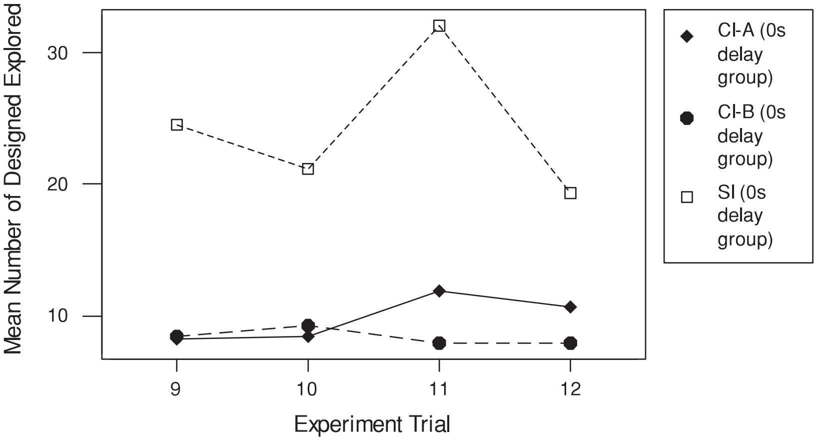


Figure 13 Mean number of designs explored for test and transfer phases of the experiment. Each subject group was trained and tested on an assigned interface with 0 s delay and then transferred to the same interface with 1.5 s delay.

face type was significant, $F(2,27) = 9.79$, $p < .001$. Next, we analyzed subject performance under reverse-delay conditions in the opposite order (see Figure 14). Once again, only interface type was significant, $F(2,27) = 6.04$, $p < .007$. Results from analysis of reverse-delay data suggest that subjects using SI consistently sampled more designs regardless of the task load.

3.3. Workload Hypothesis Testing

To test the Workload hypothesis, we focused our analysis on subjects' perceived workload as measured by the NASA-TLX. Specifically, we used the subjects' NASA-TLX

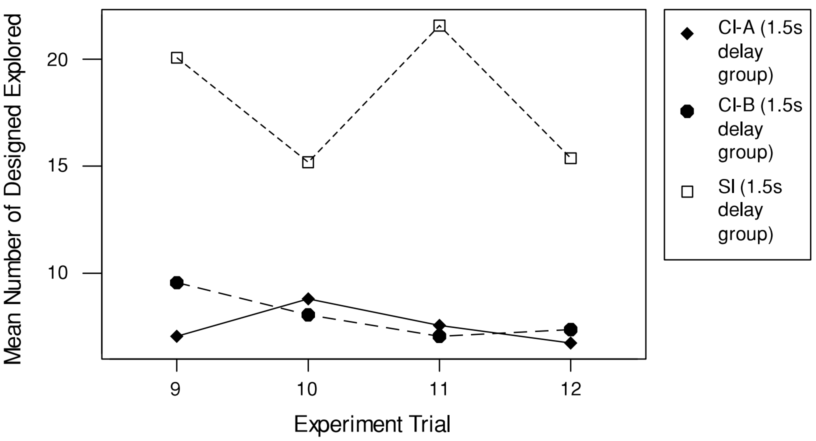


Figure 14 Mean number of designs explored for test and transfer phases of the experiment. Each subject group was trained and tested on an assigned interface with 1.5 s delay and then transferred to the same interface with 0 s delay.

TABLE 3. ANOVA Using NASA-TLX Overall Workload Score and Subscale Scores as Dependent Variables

	Factors and interactions		
Dependent variable	Interface type [$F(2,53)$]	Delay group [$F(1,53)$]	Interface × delay [$F(2,53)$]
Overall score	1.148, $p < .325$	11.51, $p < .001$	.464, $p < .631$
Mental demand	.785, $p < .461$	11.2, $p < .002$	.834, $p < .440$
Physical demand	.283, $p < .755$	14.56, $p < .000$	1.08, $p < .346$
Temporal demand	4.67, $p < .014$	9.86, $p < .003$	.526, $p < .594$
Own performance	.882, $p < .420$	2.31, $p < .135$	.358, $p < .701$
Effort	.727, $p < .488$	5.59, $p < .022$	.428, $p < .654$
Frustration	.507, $p < .605$	3.32, $p < .074$	.174, $p < .841$

Overall Workload score and subscale scores as dependent variables to examine the between-subject factor effects. A series of ANOVAs were conducted using the Overall Workload, Mental Demand, Physical Demand, Temporal Demand, Own Performance, Effort, and Frustration as dependent variables. Time delay (0.0 or 1.5) and interface type (SI, CI-A, or CI-B) were used as between-subjects factors and trial (9 and 10) as the within-subject factor. A summary of ANOVA results is given in Table 3.

The results suggest that perceived workload is affected most by the task load. Perhaps more importantly, Overall Workload is not significantly affected by the type of interface used. In fact, the only significant impact of interface type on perceived workload was found in the Temporal Demand subscale score (refer to Figure 15).

In addition to testing the Workload hypothesis, we were also interested in whether perceived workload is a good predictor of actual task performance. Linear regression analyses were conducted to evaluate the prediction of percent error and completion time from Overall Workload. The regression equation for predicting percent error is

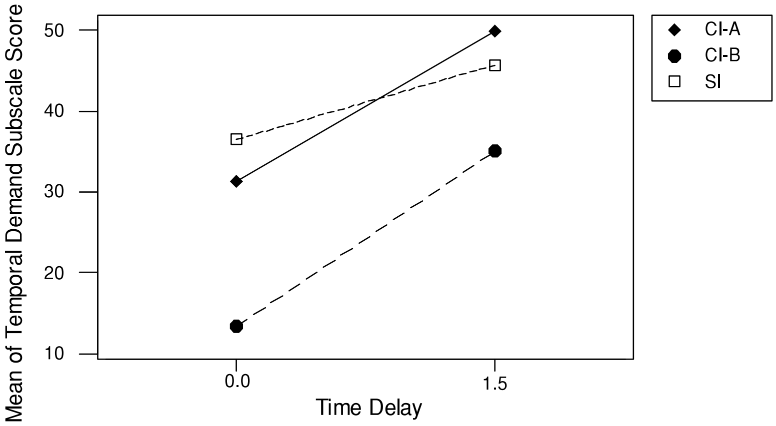


Figure 15 Mean NASA-TLX Temporal Demand subscale score.

$$\textit{Percent error} = .183\ \textit{Overall Workload} + .601 \tag{2}$$

The correlation between the Overall Workload and the percent error was .305. Approximately 10% of the variance of the dependent variable was accounted for by its linear relationship with Overall Workload. Therefore, we confirm that Overall Workload moderately predicts Percent Error scores. The regression equation for predicting completion time is

$$\textit{Completion Time} = .008\ \textit{Overall Workload} + 1.606 \tag{3}$$

The 95% confidence interval for the slope, $-.003$ to .20, indicates that the result is not significant. Thus, Overall Workload is not an adequate predictor of completion time.

4. DISCUSSION AND CONCLUSION

This study has contributed toward bridging the gulfs of execution and evaluation in computer-based engineering design. To accompany the advancement of computer-based design tools, user interfaces must facilitate human–computer interaction that is integrative, visual, and fast (National Research Council, 1998) to produce effective design solutions. We propose the use of the proximity compatibility principle (PCP) as a guide to building interfaces that deliver information required by engineers to find desired design responses. We also propose the use of the control-display compatibility principle (CDCP) to inform the layout and form of graphical objects so that engineers can effectively control the user interface. In general, this study was able to confirm predicted performance differences between SI and CI interfaces with respect to both the PCP and the CDCP.

Our PCP hypothesis stipulated that performance on the design task, which is a highly integrated task, would be better facilitated by displays that are configural instead of separable. This hypothesis was confirmed as we found that task completion time was significantly decreased under the CI-B condition without sacrificing accuracy. Moreover, the CI-B proved to be a robust interface as we found the same significant effect for subjects who transferred from a normal to an elevated task load. The advantages of the CI-B were not universal, however, as we found that percent error from the desired solution increased significantly under the CI-B condition when transferring from elevated to normal task load conditions.

Our CDCP hypothesis proposed that the CI-B and CI-A interfaces, which are higher in control-display compatibility (see Table 2) will enable more efficient performance on the design task. Our hypothesis was confirmed by the finding that the number of designs explored significantly increased under the SI condition. This finding indicates that configural interfaces enabled more efficient performance than the separable interface through fewer points explored and higher performance scores. Moreover, the benefit of configural displays was found to persist under normal or elevated task load. Note that the CDCP hypothesis does not directly address sequential control versus simultaneous control of multiple parameters—the difference between CI-A and CI-B. However, since the superiority of CI-B over other interfaces (in terms of latency) has already been shown, we indirectly confirm the CDCP hypothesis because the difference between CI-A and CI-B lies in the perceptual availability and control of input variables.

Our Workload hypothesis suggested an inverse relationship between designer-perceived workload and the PCP and CDCP. Therefore, we expected to find a significant difference in workload ratings between configural and separable interfaces. However, the Overall Workload score was not significantly affected by interface type. Moreover, interface type was a significant effect for only one of the six NASA-TLX subscale scores. This significant effect was consistent with our hypothesis as time pressure felt by subjects was significantly lower in the CI-B condition. Nevertheless, we found that the NASA-TLX is a moderate predictor of performance in the I-beam design task, and its consideration should be secondary to performance accuracy and latency.

One implication of developing display design principles is to inform the design of automatic graphic design and presentation tools like BOZ (Casner, 1991). One can envision the creation of a tool that forms compositions of existing integral and separable designs to arrive at displays that customizes displays to support the operators in computer-based engineering design decision-making tasks.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant No. DMI-0084918. Any opinions, findings, and conclusions or recommendations presented in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Barnett, B.J., & Wickens, C.D. (1988). Display proximity in multicue information integration: The benefit of boxes. Human Factors, 30(1), 15–24.

Barron, K., Simpson, T.W., Rothrock, L., Frecker, M., Barton, R.R., & Ligetti, C. (Submitted). Graphical user interfaces for engineering design: Impact of response delay and training on user performance. Research in Engineering Design.

Barton, R.R. (1998, December). Simulation metamodels. Paper presented at the 1998 Winter Simulation Conference, Washington, DC.

Beard, D.V., & Walker, J.Q. (1990). Navigational techniques to improve the display of large two-dimensional spaces. Behaviour & Information Technology, 9(6), 451–466.

Bennett, K.B., Nagy, A.L., & Flach, J.M. (1997). Visual displays. In G. Salvendy (Ed.), Handbook of human factors and ergonomics (pp. 659–696). New York: Wiley.

Bennett, K.B., Toms, M.L., & Woods, D.D. (1993). Emergent features and graphical elements: Designing more effective configural displays. Human Factors, 35(1), 71–97.

Bullinger, H.J., Kern, P., & Braun, M. (1997). Controls. In G. Salvendy (Ed.), Handbook of human factors and ergonomics (pp. 697–728). New York: Wiley.

Carswell, C.M., & Wickens, C.D. (1990). The perceptual integration of graphical attributes: Configurality, stimulus homogeneity, and object integration. Perception and Psychophysics, 47, 157–168.

Casner, S.M. (1991). A task-analytic approach to the automated design of graphic presentations. ACM Transactions on Graphics, 10(2), 111–151.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Gerace, J., & Gallimore, J.J. (2001). Evaluation of visual display techniques for assembly sequence planning. Human Factors and Ergonomics in Manufacturing, 11(3), 213–231.

Gibson, J.J. (1986). The ecological approach to visual perception. Hillsdale, NJ: Erlbaum.

Goodman, T., & Spence, R. (1978). The effect of system response time on interactive computer-aided design. Computer Graphics, 12, 100–104.

Haftka, R., & Gürdal, Z. (1992). Elements of structural optimization (3rd ed.). Boston, MA: Kluwer.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139–183). Amsterdam: North Holland.

Hirschi, N.W., & Frey, D.D. (2002). Cognition and complexity: An experiment on the effect of coupling in parameter design. Research in Engineering Design, 13, 123–131.

Hutchins, E.L., Hollan, J.D., & Norman, D.A. (1986). Direct manipulation interfaces. In D.A. Norman & S.W. Draper (Eds.), User centered system design (pp. 87–124). Hillsdale, NJ: Erlbaum.

Jones, C.V. (1994). Visualization and optimization. ORSA Journal of Computing, 6(3), 221–257.

Karwowski, W. (2000). Symvatology: The science of an artifact–human compatibility. Theoretical Issues in Ergonomics Science, 1(1), 76–91.

Ligetti, C., Simpson, T.W., Frecker, M., Barton, R.R., & Stump, G. (2003). Assessing the impact of graphical design interfaces on design efficiency and effectiveness. ASME Journal of Computing and Information Science in Engineering, 3(2), 144–154.

Messac, A., & Chen, X. (2000). Visualizing the optimization process in real-time using physical programming. Engineering Optimization, 32(6), 721–747.

National Research Council. (1998). Visionary manufacturing challenges for 2020 (Committee on Visionary Manufacturing Challenges Report). Washington, DC: National Academy Press.

Norman, D.A. (1991). Cognitive artifacts. In J.M. Carroll (Ed.), Designing interaction: Psychology at the human–computer interface (pp. 17–38). New York: Cambridge University Press.

Petre, M., & Green, T.R.G. (1992). Requirements of graphical notations for professional users: Electronics CAD systems as a case study. Le Travail Humain, 55(1), 47–70.

Pomerantz, J.R. (1986). Visual form perception: An overview. In H.C. Nusbaum & E.C. Schwab (Eds.), Pattern recognition by humans and machines (Vol. 2, pp. 1–30). Orlando, FL: Academic Press.

Sanderson, P.M., Flach, J.M., Buttigieg, M.A., & Casey, E.J. (1989). Object displays do not always support better integrated task performance. Human Factors, 31(2), 183–198.

Simpson, T.W., Peplinski, J., Koch, P.N., & Allen, J.K. (2001). Metamodels for computer-based engineering design: Survey and recommendations. Engineering with Computers, 17(2), 129–150.

Tullis, T.S. (1988). Screen design. In M.G. Helander (Ed.), Handbook of human–computer interaction (pp. 377–411). Amsterdam: Elsevier.

Vicente, K.J., & Rasmussen, J.R. (1992). Ecological interface design: Theoretical foundations. IEEE Transactions on Systems, Man, and Cybernetics, 22(4), 589–606.

Wickens, C.D. (1992). Engineering psychology and human performance (2nd ed.). New York: HarperCollins Inc.

Wickens, C.D., & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. Human Factors, 37(3), 473–494.

Wierwille, W.W., & Eggemeier, F.T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. Human Factors, 35, 263–282.