15

Evaluating User Performance to Experienced Level and Beyond

Jackie Moyes

Information Technology Institute
11 Science Park Road
Singapore Science Park II
SINGAPORE 117685
moyes@iti.gov.sg

ABSTRACT

Studies focusing on two iconic attributes, form and relative position, suggested subjects rely on form to recognise a particular icon when the form is representational. However, when form is abstract users switch from attempting to use form, to using position. Results had suggested that some degree of positional learning still occurred for subjects using representational form. However, a more in depth study of this suggestion was impossible since the standard performance measures of error rate and reaction time settled quickly into an optimal level of performance. To gain better insight into what was occurring, cursor movements were tracked, allowing pre-emptive move data to be captured. The showed that, irrespective of the type of form presented, subject learned and eventually switched to relying on position to identify the icon. More importantly, results suggest that this learning continued after the traditional measures of reaction time and error rate had reached what Jordan et al. (1991) termed experienced user performance level, or asymptote. The suggestion therefore is that traditional performance measures are perhaps not sensitive enough to measure learning at the experienced end of the learning curve, and that interface designers should thus consider ways in which to capture this potentially enlightening information which, as yet, remains unexamined.

KEYWORDS usability evaluation, icon design, quantitative measures, pre-emptive move data

1. INTRODUCTION

A series of experiments had been conducted with the aim of trying to predict which icon attributes users were likely to rely upon to associate a particular icon with a command (Moyes 1995). This study originated from the observation that most designers stressed the importance of the icon form in communicating the icon's meaning, while ignoring other potential attributes (e.g. position, colour, size, outline shape, etc.). The theory being tested, therefore, was that initially users will rely on the attribute which was the most representational, but, over time, may switch to identifying the icon using another attribute, if this attribute offered an increased performance advantage.

This was in line with the theory established by Kaptelinin (1993) who suggested that users focus on local attributes initially (e.g. icon form) but over time switch to identifying attributes that are global, or require the user to consider the interface as a whole into consideration (e.g. the icon's relative position amongst all other icons on the screen).

Human–Computer Interaction: INTERACT'97 S. Howard, J. Hammond & G. Lindgaard (editors)
Published by Chapman & Hall

Additionally, in an attempt to formally identify when users were likely to switch between attributes, a multi-component definition of usability, as established by Jordan et al., 1991), was adopted. This definition suggested that there were three identifiable stages in the learning curve (as illustrated in Figure 1); guessability, learnability and experienced user performance (EUP). Their definitions are as follows:

- Guessability - the measure of the cost to the user involved in using an interface to perform a new task for the first time. The lower the cost, the higher the guessability (cost can be measured either in terms of time, errors, or effort).
- Learnability - the measure of the cost to the user in reaching some reasonable level of performance on a task, but excluding the special difficulties of completing the task for the first time. A highly learnable interface would be one where a task was instantly memorable once the method had been shown to the user. Conversely, interfaces which cause 'interference' with user expectations are likely to be un-learnable.
- EUP - a measure of the ability of the user to perform a task when he or she has reached a relatively steady level of performance. Again, the lower the cost, the higher EUP

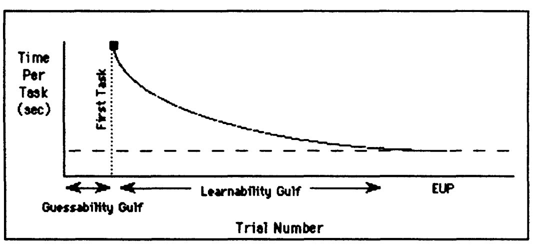

Figure 1. The learning curve with the three stages of usability illustrated.

The series of experiments conducted had focused on the attributes form and position, and had studied user performance, using these attributes, throughout the entire learning curve. It was predicted that if form was representational then users would ignore the icon's position and rely on form to identify the icon. However, if the icon form was abstract then users would initially attempt to use form to identify the icon, but since position offered a better performance advantage, in the learnability stage they would switch to position to identify the icon.

To test this theory 4 experimental conditions were created:

- Representational -> Blank
  In this condition subjects were trained to EUP using icons with representational form (see Figure 2) and fixed position. Once EUP had been achieved an unexpected change would take place and position would remain the same but icon form would be removed. The position of the icons would be marked by empty rectangles
- Abstract -> Blank
  Here subject were initially presented with icons in their abstract form (see Figure 2), and fixed position. At changeover the position of the icons would again remain the same, but the form would be removed.
- Representational -> Random
  For this condition users were presented with representational form throughout, but at changeover the position of the icons was be randomised.
- Abstract -> Random
  Users were presented with abstract form throughout, but at changeover the position of the icons was randomised.

It was predicted that if the theory was correct, subjects in conditions 1 and 4 would be affected by the changeover (indicated by significantly slower reaction times and increased errors) since the attribute that they had been relying on had been removed, thus requiring a degree of re-learning to occur. Subjects in conditions 2 and 3 should not show any significant disruption since the attribute they had been predicted to rely on had not changed.

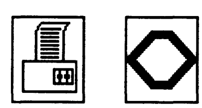

Figure 2: The print icon. One representational and one abstract version, as used in the experiment (actual size).

Results from previous experiments support this hypothesis. At changeover the biggest disruption in performance was found for the condition where representational form was removed (condition 1), followed closely by the condition where abstract form icons were randomised (condition 4). However, although there was a significant disruption, in each of these cases, recovery to performance levels similar to those achieved before changeover was extremely rapid. It was therefore unclear whether disruption at changeover was simply due to surprise rather than the need to relearn. In other words, perhaps the quick recovery rate was indicative of the fact that subjects in all conditions had actually learned both attributes to some degree. This may be acceptable for conditions 3 and 4 where subjects were expected to have initially attempted to learn icon form, but moved to using position later in the learning curve. However, it had been predicted that subjects using the representational form icons would remain using this attribute, therefore the question being posed was whether subjects in these conditions were indeed learning position.

The experiment presented here replicates the methodology used in previous experiments, but with two major differences. Firstly, the size of the icon set is increased from 2 to 16. In previous experiments only 2 icons had been presented on the interface. It was therefore assumed that perhaps an additional reason for the fast recovery rates was because subjects found learning the new attribute extremely easy when there was only a two way choice involved. Increasing the number of icons displayed to 16 was hoped to make the interface more complex, thus increasing the amount of learning time involved (assuming no learning of the attribute had taken place prior to the changeover).

The second, and more interesting change was that cursor movement was recorded. A standard trial consisted of presenting subjects with a screen showing only the command they were required to find the appropriate icon for, and an OK button. Subjects read the command and then clicked on the button to move onto the interface. It was possible therefore to record the movement of the cursor between the time when the user clicked the button until the time when the interface appeared. If the trajectory of the cursor appeared to be moving towards the position of the correct icon before the interface was actually visible, then this was marked as a correct pre-emptive move, and assumed to be indicative that the subject had learned that icon's position.

This new quantitative measure could then be compared between the 4 conditions to give a better idea of whether subjects in the representational conditions were indeed learning position, and if so, to the same extent as users in the abstract form conditions.

2. METHODOLOGY

2.1 Apparatus

The experimental sessions were performed using a Macintosh Quadra (system 7.1.2).

2.2 Design

The dependent variables were:

- Time taken in each trial
- Number of errors
- Number of pre-emptive moves

The independent variables were:

Icon form	**2 levels** Representational Abstract
Position	**2 levels** Consistent Random

Subjects were grouped into one of 4 possible conditions:

- representational -> blank
- abstract -> blank
- representational -> random
- abstract -> random

Sixteen icons were used for each set.

Changeover only occurred after users had been recognised as EUP; namely by the time reaction time and error rate had reached optimal, steady levels. For this experiment performance levels were predicted to level by block 20 (20 x 16 trials), therefore changeover occurred at block 21.

2.3 Subjects

Sixteen subjects were used (four per group). Thirteen subjects were male and three were female. The mean age was 26.

2.4 Procedure

Subjects were initially presented with a block of practice trials to familiarise themselves with the experimental task. Although the task itself was identical to that performed in the experiment a mock experimental interface was created where the icons in each of the positions were replaced by numbers (e.g. 1), and the commands were in textual format (e.g. Number One).

The trial commenced when the command screen, showing the command and OK button situated in the centre of the screen. Once the user clicked on the button the interface card opened, displaying an icon set. The icons were positioned in an equi-distant circular format. The subject was required to respond by clicking the mouse over the icon he or she considered to be appropriate for the command. If a correct response were made the set disappeared and approximately 0.5 seconds later the next trial commenced. An erroneous response was signalled by a high tone and a penalty was induced. A dialogue box appeared and subjects were informed that they had made an error and would be shown the correct icon. The subject clicked the OK button, the dialogue box disappeared and the correct icon was indicated, after which the next trial commenced.

2.4.1 Details Recorded

For every trial eight pieces of data were captured. These were:

1. The errors. Each icon had a number, and the number of the correct icon was stored, followed by the number(s) of any incorrect icons that were selected.
2. The time that the command and the OK button appeared.
3. The time that the subject clicked on the button to move to the interface.
4. The (x,y) position of the pointer on the screen as the screen closed. This showed whether or not the mouse was outside the button area, and in what direction it was travelling. This data was collected from experiment 2 onwards.
5. The time that the screen closed.
6. The (x,y) position of the cursor as the interface screen opened.
7. The time that the first icon was selected (and whether it was correct or incorrect).
8. The time that the interface screen closed.

A typical reaction time was measured from the time that the command appeared until the user had selected the correct icon, or been informed of its position, and the interface card had closed.

3. RESULTS

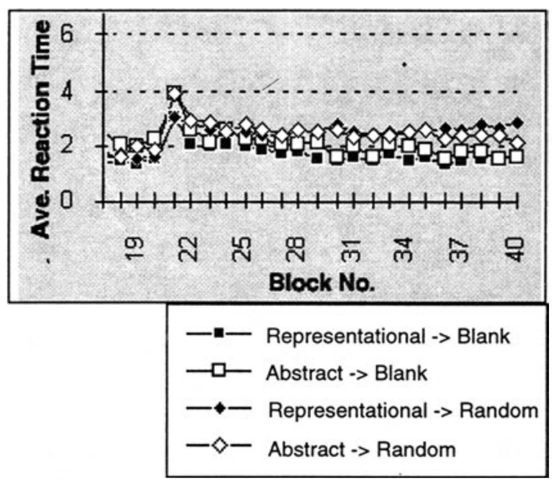

Figure 3. Graph of the mean reaction times across blocks 1 to 40.

The main aims of this experiment were:

- To evaluate whether the recorded disruption was due to surprise or learning.
- To discover when subjects began to learn position and whether subjects in all conditions learned it or just those using icons with abstract icon forms.

Interpreting the results from the traditional dependent variables, reaction time (Figure 3) and error rate (Figure 4), it would appear that the disruption not due to surprise, since in both cases, the disruption lasts beyond 1 or 2 blocks of trials. The results, however, do appear to indicate that some degree of position learning by subjects in the representational conditions, prior to the changeover, had indeed occurred. Even using larger icon sets, the disruption at changeover, in comparison with the performance levels at the start of the test is, for all conditions, minimal. Additionally, any evidence of disruption, after block 25 is almost imperceptible.

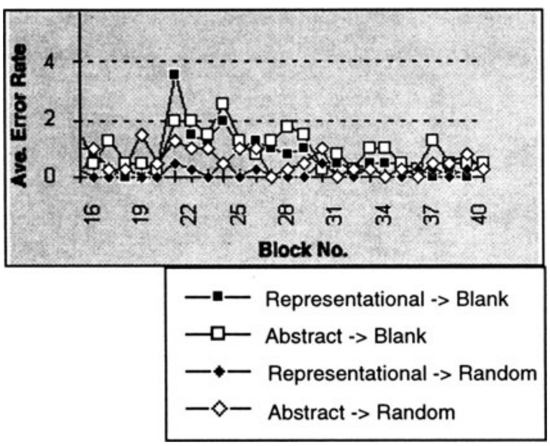

Figure 4. Graph of mean errors across 40 blocks of trials

Figure 5, displaying the graph of pre-emptive moves however tells a far more enlightening story. A pre-emptive move was recorded whenever the trajectory of the cursor from the OK button on the command screen to the correct icon on the interface card could be plotted as a direct line. If pre-emptive moves are indeed taken as evidence for position knowledge, then the graph

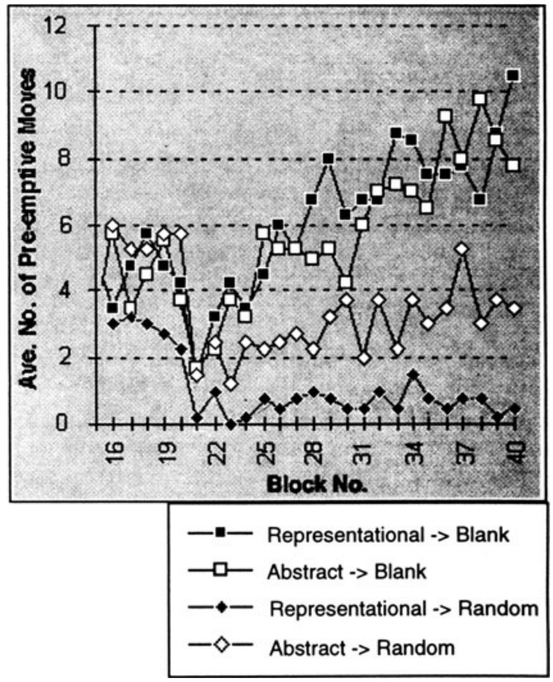

Figure 5. Graph of the mean number of correct pre-emptive moves made in each condition across 40 blocks of trials (16 being the maximum number of moves possible).

suggests that position is learned by subjects in all 4 conditions. There appears to be a gradual learning curve that commences at block 1 and increases steadily until the disruption at block 21. The 2 conditions where form is removed quickly recover to an almost identical learning curve shortly after the changeover. Where position is randomised there is a difference between the groups. In both cases there is a significant drop in the degree of pre-emptive moves made. However, the group using the representational form appear to give up on position almost entirely. However, there is an indication that some learning of position does occur in the abstract condition. This is likely to be an effect of the experimental design, since the position of each icon was randomised to appear in 1 of 2 possible positions. This result, therefore, suggests that people using abstract icons found it easier to try to learn both icon positions, and make a guess as to which one the icon would actually be located in, rather than rely upon the icon's form. In other words, they would rather learn 32 possible positions than 16 iconic forms.

By the end of 40 blocks of trials subjects in the conditions where position remains constant appear to be making pre-emptive moves for an average of 10 out of every 16 trials. This result is extremely important since it suggests that learning has not ended, even though the traditional measures of reaction time and error rate have reached optimal levels.

4. DISCUSSION

Therefore, results appear to suggest that it was not surprise that caused the disruption, since the disruption lasted longer than one or two trials. However, the disruption in almost all cases tended to be shorter than was predicted if the user was learning a completely new attribute. This suggests that subjects had learnt the attribute, considered in the initial hypothesis to be the recessive attribute, and so were able to return to a similar level of performance in a short period of time.

The pre-emptive move data convincingly supports the idea that position is learned by subjects in all conditions. The graph in Figure 4 suggests that there is a gradual learning curve commencing from block number one for all conditions. This confirms the assumption that people learn form first and then move on to learn position. However, the speed of learning position as opposed to form in this experiment could be influenced by the obvious performance advantage offered by knowing icon position. This is an extremely important point in interface design since it suggests that the interface context (e.g. the amount of information present on the interface) is going to have a strong influence on which attributes are used.

The graph also suggests that by block 40 users have not yet reached an optimal state in the number of pre-emptive moves they will make. It is possible that the curve will plateau only when the user is able to make pre-emptive moves for all icons in the set. Again this result is probably influenced by the high performance advantage offered by learning position. The interesting point to note, however, is that both reaction times and error rates have stabilised around block 30, yet the pre-emptive move data suggests that learning is still occurring. So, although learning appears to continue, reaction times and error rates fail to detect it. Therefore, other measures need to be considered if learning beyond what Jordan et. al (1992) define as EUP is to be detected and analysed.

5. CONCLUSION

This latter point is extremely important. It appears that an attribute considered to be important in the guessability stage of performance is not necessarily the attribute used throughout the rest of the usability curve. There is a distinct contrast between the "effort after meaning" made by the user in the guessability stage, and demand for a highly discriminable attribute with a small cognitive load in the learnability and EUP stages. Therefore as the experiments have suggested, icons which contain attributes which appear to offer no meaning to the user in the guessability stage perform as well as icons possessing these meaningful attributes at EUP. Although icon-oriented interfaces tend to be aimed at the computer novice, by restricting study of icon design to how to attain meaningful icons, researchers are ignoring the potential design issues apparent to the majority of the usability curve.

However results have shown that learning and fine tuning of procedures to recognise particular icons with minimum effort and cognitive load continue to occur after experienced user performance level has been achieved using the classical performance measures of time and error. This may suggest that Jordan et al.'s (1991) model of usability may be too simplistic. With the ability to capture continued learning through measures such as the number of pre-emptive moves a more detailed picture may be achieved.

There are also many questions raised by considering performance throughout the learning curve. As suggested, we could ask does EUP exist? Do users ever reach a plateau, or do they continue fine tuning their performance? Do users fine tune their performance to such an extent and then switch off to any changes in the environment, or do they continue to monitor change? How can we measure EUP beyond the stage where reaction time and error rates reach optimal performance levels? Pre-emptive move data may be an important step towards this, however, this methodology is limited to situations where learning of position is being measured. It does, however, indicate that more creative usability measures need to be developed, perhaps not only for icon design, but for screen design in general.

6. ACKNOWLEDGEMENTS

Thanks to Phil Gray, Steve Draper and Paddy O'Donnell for their advice, practical suggestions and encouragement during this period of research.

6. REFERENCES

Jordan, P W, Draper, S W, MacFarlane, K K & McNulty, S-A. (1991) Guessability, Learnability, and Experienced User Performance. In *People and Computers VI,* Diaper, D. & Hamond, N. (eds.), Cambridge University Press.

Jordan, P.W. & O'Donnell, P.J. (1992) Quantifying guessability, learnability and experienced user performance. *Contemporary Ergonomics 1992*, Lovesey, E.J. (ed.), London: Taylor & Francis, pp. 404-410.

Kaptelinin, V. (1993) Icon Recognition in Menu Selection: The Effect Of Practice. In *INTERCHI'93 Adjunct Proceedings*, Ashlund, S. et al. (eds.), ACM Press.

Moyes, J. (1995) Putting Icons in Context: The influence of Contextual Information on the Usability of Icons. *Ph.D. Thesis*, University Of Glasgow.