Int. J. Human-Computer Studies (2001) **55,** 485–501
doi:10.1006/ijhc.2001.0478
Available online at http://www.idealibrary.com on IDEAL®

Prototyping for usability of new technology

ROGER R. HALL
School of Safety Science, UNSW Sydney 2052, Australia. email: r.r.hall@unsw.edu.au

(Received 15 November 2000, and accepted in revised form 29 April 2001)

This paper argues that it is possible to gain good design information from low-cost user trials of low-fidelity prototypes early in the design process, and that simple prototyping is a valuable tool in the user-centred design of new technology especially "smart" consumer products. The value of that design information depends on the stage of the design process at which user testing is carried out and the associated level of realism or fidelity of the prototype. The first stages involve testing simple prototypes which examine the cognitive, or information processing, needs of the user, followed by higher-fidelity prototypes which examine the physical (visual, auditory and tactile) needs of the user.

The results of four studies are discussed to illustrate: the extent and nature of the design information gathered, the relative merits of varying the fidelity of the prototypes, and the benefits and costs associated with using different levels of fidelity of prototypes in a user-centred approach to design. Finally, and based on that discussion, an appropriate and practical design strategy is suggested.

KEYWORDS: user-centred design; usability; prototypes; value; efficiency.

1. Introduction

It is well accepted that people have problems using new technology such as consumer products, systems and software. There is some evidence to show that poor design can cost companies money directly through attending to user calls for service and complaints, and through returned goods. Poor design can also indirectly cost companies through reduced sales because of poor consumer acceptance and poor product image, and through the associated follow-on effects of consumer perceptions of the company itself. This is becoming especially so with the advent of embedded computer processors in the so-called "smart" domestic consumer products.

Such use of embedded computer chips facilitates an increase in the functionality of a product but often to the detriment of its usability. Norman (1988) calls this phenomenon *creeping featurism* and suggests that the complexity of use increases as the square of the increase in functionality. That people do have problems with new technology is not a new phenomenon but has been around for many years with such technology as video cassette recorders (VCRs), automatic teller machines (ATMs), microwave ovens and ticket vending machines (Evans & Moore, 1991; Thimbleby, 1991; Adams & Hall, 1992; Maguire, Butters & Mcknight, 1994).

More recently, it has been evidenced by examples of technology becoming smaller and smaller but at the same time with more programmable functions and other

1071-5819/01/000000 + 13 $35.00/0

market-driven features. Mobile telephones, for example, are becoming very small but now include features such as programmable ring tones, a calculator, games, the capability to send short text messages, and even internet access for email correspondence and world wide web browsing.

Size reduction usually imposes physical constraints on the design of the user–technology interface. This can lead to problems in use through small, multi-modal buttons and, at the same time, only limited visual feedback in small, and often, non-backlit LCD displays. However, other examples of consumer technology which have sufficiently large, physical interfaces are not always immune from usability problems. For example, some public transport, automatic ticket vending machines have been shown to be very poorly designed (Adams & Hall, 1992).

2. User-centred design

To avoid usability problems designers should employ an ergonomics or user-centred approach in their design work (cf. Norman & Draper, 1986; Norman, 1988, 1998; Shneiderman, 1998). A user-centred approach involves knowing who the users will be, their capabilities, needs and expectations, their goals and the tasks required to achieve those goals, and the physical and social environments in which users have to achieve those goals. It involves processes of participatory design, user testing and iterative design (Gould & Lewis, 1985; Shackel, 1986; Gould, 1995). Participatory design implies involving users as participants in the design team while user testing involves users as subjects in the testing of design concepts or mock-ups from the very beginning of the design process. Iterative design implies a cycle of design, test and redesign, retest, etc. Gould and Lewis (1985) suggest that designers may believe "that iteration is just expensive fine tuning" but they argue that "with the current state of understanding about user interface design, it [iterative design] is the only way to ensure excellent systems". These general processes have been more or less formalized in International Standard 13407 (ISO, 1999) which also confirms the need to determine first the appropriate allocation of functions between user and technology.

However Eason (1992) describes three forms of user-centred design depending on the "type" of design being developed.

1. Generic, i.e. design for the user where the role of the user is as a subject in user testing of the design, e.g. consumer products or general software.
2. Bespoke, i.e. design by the user where the role of the user is as a participant in the design process, e.g. a software application for a company's specific needs.
3. Customizable, i.e. design adapted by the user where the role of the user is to control how the design interface looks and/or operates for them, e.g. individualizing the interface to personal computer operating systems.

He also describes a generic form of user-centred design the aim of which is as follows.

- First, to help users articulate their needs as part of the overall product *requirements* which are a function of both the context in which the product is used as well as of the particular tasks to be performed.

- Second, to formulate various *options* which have to meet both technical and/or social constraints.
- Thirdly, to *evaluate* the degree to which the options meet the requirements.

Eason states that external experts may be needed to help identify the options but that users "own" the requirements. Also, users will "need help to evaluate whether a given option will meet the functional requirements" as well as to assess its impact on usability and acceptability. This generic form also facilitates the iteration process as once the *evaluation* stage has been done, new options can be developed, and evaluated again, and again if necessary.

Hall (1997) has extended Eason's model by suggesting that by first using low-fidelity mock-ups of the initial design *options*, the results of user *evaluations* can be fed back into a review of the user needs parts of the product *requirements*. Then, based on any revision of those requirements, extra fidelity can be added to the (new) design *options* which can then be *evaluated* by users to provide further feedback into a second review of the requirements and, if required, a third iteration of the design process.

It may be argued by some that this already occurs; however, the author contends that, in practice, the (user) requirements are rarely changed because they have already been signed off by the client and are therefore deemed not to need reexamination, even in the light of the results of user testing. Rather, the user testing involves determining which of various alternative designs best meet the already agreed requirements. Further, Ward (1994) suggests that there need to be studies which demonstrate that user testing during the design process will provide information which will result in a better design and that such user trials are within the resources and capabilities of (industrial) designers. User testing involves some form of mock-up or prototype of the design in sufficient fidelity to evaluate the design itself or a specific aspect or module of the design currently under consideration.

3. Evaluation and the design process

There are various methods available to evaluate the usability of a design. The most appropriate method depends on the type and complexity of the product or system being designed, and the stage of the design process at which the "evaluation" is to be done. Preece, Rogers, Sharp, Benyon, Holland and Carey (1994) and Dix, Finlay, Abowd and Beale (1993) both give an overview of various methods suitable for designing a human–computer interface, particularly software interfaces. The method of Preece *et al.* include *observing and monitoring* users' performance, *collecting users' opinions* of their use of the product, running *experiments and benchmarking* to gather scientifically based quantitative data on the human–computer interaction, and *interpretive* and *predictive* evaluations. They also contrast the purpose of *formative* and *summative* evaluations and highlight the necessary link between *formative* evaluation and design: "[formative] evaluation meshes closely with design and guides the design by providing feedback".

Evaluation can take place at many stages during the design process. A simple form of the *design process* is given by Baecker and Buxton (1987, p. 503) as follows.

1. Collect information.
2. Produce a design.

3. Build a prototype.
4. Evaluate the system
5. (a) If not acceptable, revise 2 and/or 3 as appropriate and (b) if acceptable, deliver the system.

Cushman and Rosenberg (1991) give a more detailed *product development process* which includes the following phases.

1. Product planning.
2. Design.
3. Testing and verification.
4. Production.
5. Marketing and evaluation.

Some form of evaluation (product testing) and iteration (redesign and evaluate) is included in both models of the design process. Cushman and Rosenberg (1991) state that "product testing begins during the design phase of product development with an evaluation of the first mock-ups and engineering models, and it concludes with verification tests of advanced prototypes at field sites". They say that evaluations can be performed by ergonomics or human factors experts but "in some cases it is appropriate to have representative users participate in the evaluations in order to get early feedback concerning usability".

Preece *et al.* (1994) go further by defining evaluation, within the context of designing the interaction between humans and computers, as particularly being "concerned with gathering data about the usability of a design or product by a specified group of users for a particular activity within a specified environment or work context". What both sets of authors are saying is that, as part of the design process, it is necessary to evaluate designs iteratively using representative users performing appropriate tasks or work, within specific environments or work contexts. This process of user testing of design prototypes is generally called prototyping.

4. Prototyping

Prototyping is not new. As long ago as 1959, Shackel investigated the design problem of "positioning 24 potentiometers and 24 associated switches in the minimum panel size consistent with easy location and control by the operator", using alternative layouts (i.e. designs) drawn to actual size on paper (Shackel, 1959). Seminara and Gerrie (1966) successfully investigated alternative control–display panel mock-ups using magnetized panel elements, and also tested three-dimensional mock-ups of work station designs to be incorporated within a mobile lunar laboratory.

Early use of prototyping concentrated on the physical aspects of design (layouts, workstation dimensions, etc.) but as computer technology became more widely available, prototypes could be easily and quickly developed on computer screens (Metz, Richardson & Nasiruddin, 1997; Wright & Monk, 1989; Virzi, Penn, Tullis & Greene, 1990; Blatt & Knutson, 1994; Hall, Zinser & Keller, 1999). This has facilitated the ability to conduct the so-called rapid prototyping, either in the form of software interface story boards with hypermedia tools, or computer-controlled production of solid form models of products which allow the look and feel attributes of a product to be evaluated.

Prototyping can take many forms and be carried out for many purposes. Jordan (1998) states that "there are a number of different prototyping options, of differing degrees of realism and sophistication, which can be used in the design/evaluation cycle." He gives examples beginning with a verbal or written description of the form and functionality of the proposed product; to visual prototypes (drawings on paper or screen); to some physical representation, or model, of the product; to screen-based interactive prototypes; and ending with fully working prototypes.

Stanton and Young (1999), in examining suitable ergonomics methods to improve product design, make a distinction between the types of prototypes according to the principal stages of the design process as follows.

1. Concept.
2. Design.
3. Analytical prototype.
4. Structural prototype.
5. Operational prototype.

Here an analytical prototype corresponds to a computer-aided design and the structural prototype corresponds to a "hard-built" prototype. The operational prototype refers to the design ready for commissioning. This is somewhat similar to Meister's (1990) design process which refers to mock-up testing being performed during the *planning*, *preliminary design* and *detailed design* stages, and operational testing of prototype systems during the *production and deployment* stage.

While Meister (1990) focuses solely on performance measures of system effectiveness, Stanton and Young (1999) focus on the standard usability evaluation methods and state that "some methods are better to use on potential end-users of the product … and you should choose your method according to the output you require: errors, performance times, usability or design".

Nielsen (1993) describes an 11-stage process which he calls the *usability engineering lifecycle* beginning with "1. Know the user" to "11. Collect feedback from field use" and including participatory design, prototyping and iterative design stages. He also puts forward the idea of *horizontal* and *vertical* prototypes where a horizontal prototype retains all the features of the system but eliminates depth of functionality while a vertical prototype provides full functionality for a few of the features. The full system is therefore defined as having all the features with complete functionality of those features. Neilsen's notion of a horizontal prototype describes what earlier authors have evaluated (e.g. Shackel, 1959; Seminara and Gerrie, 1966) and has proven to be a very useful approach to determining the optimum layout of controls and displays, rather than the physical attributes (e.g. tactile feedback) which would come with evaluation of a vertical prototype.

In 1991 Ward, then a practising industrial designer, investigated whether the involvement of users early in the design process could lead to a better design for a programmable, domestic lighting controller being built by a client (Ward, 1994). He reported on the value of using a low-fidelity, cardboard mock-up of the controller interface to identify potential problems in its use. In other words he was evaluating a horizontal prototype

FIGURE 1. Ward's cardboard and foam mock-up of the original design of the domestic lighting controller.

because there was no functionality to the "switches" on the panel. Some functionality to the "LCD display" was provided by "slip-in" cards on which system prompts were written and which were visible through a rectangular cut-out in the cardboard panel (see Figure 1).

Due to the limited functionality Ward did not believe that measuring task times would be valid or helpful. Instead he believed that by understanding the problems which users had in interacting with the panel any deficiencies in the design would be revealed. He defined a problem as "any feature (or omission) in the product or instructions that was identified as contributing to the following.

1. An error—any action that would have produced a system response other than that intended.
2. Confusion—where the subject could not continue without assistance, even after referring to the instructions.
3. Irritation or dislike—where the subject commented that a particular feature was difficult to use or not the way they would have preferred—even if this did not actually result in an error".

He took a representative sample of 10 users who tested the mock-up by carrying out some routine tasks associated with the controller. He was able to identify 28 different problems and, based on that feedback, he was able to revise the design, test it, and in doing so was able to eliminate 20 of the original problems and significantly reduce the average number of problems experienced by 70%; however, as can often happen, a new problem was introduced into the revised design!

Ward chose a very low-fidelity prototype, a cardboard and foam mock-up, because he wanted to show fellow designers that such user testing of design ideas or models could provide valuable design information within reasonable time and cost. He estimated that it took him half a day to make the cardboard mock-up and only several days to test the original design.

5. Fidelity in prototyping

Prototypes can take various forms from renderings and diagrams on paper, cardboard and/or foam representations (usually with limited or non-functioning controls and displays), touch screen simulations of interface panels etc., three-dimensional simulators, to virtual reality environments. The attribute which distinguishes these prototypes is their relative fidelity or faithfulness in reproducing the characteristics of the finished product. The degree of fidelity would represent some combination of Nielsen's (1993) attributes of both horizontal and vertical prototypes.

There have been numerous studies using differing levels of fidelity, for many different products. A non-exhaustive list, indicating the type of prototype (in increasing order of fidelity), the usual format (two- or three-dimensional) and degree of possible interaction, and relevant studies, is given in Table 1.

Säde (1999) presents a different form of classification of prototypes which he refers to as models but calls *representations of product concepts*. It includes "use scenarios" for idea generation which is the first stage in the designing process. Säde's classification is consistent with that of Table 1 although different terms are used. For example some forms of "use scenarios" could be included in Type 1 "renderings ... ", and the "drawing

TABLE 1
Type of prototype, usual format and degree of interaction and relevant studies

Type of prototype	Usual format and degree of interaction	Relevant studies
1. Renderings or diagrams of design concepts	2D form and layouts with little interaction	Shackel (1959), Rooden (1999)
2. Cardboard and/or foam mock-ups	2D layouts and some 3D form with limited interaction	Ward (1994), Hall (1999), Rooden (1999)
3. VDU-based prototypes, 2D simulations	2D form with some interaction	Metz *et al*. (1987), Kenney (1994), Hall *et al*. (1999), Hall (1999)
4. Solid form prototypes	3D form with some interaction	Kahmann and Henze (1999)
5. 3D simulations	3D form with significant interaction	de Vries and Johnson (1999), Porter and Porter (1999)
6. Virtual reality environments	3D form with significant interaction	Karwowski, Chase, Gaddie, Lee and Jang (1997)
7. The real product or system	Actual form with total interaction	Rooden (1999)

and storyboarding" and "paper prototypes" representations could also be included in Types 1, 2 or 3 depending on the design context.

Also, Säde (1999) argues that "iterative modelling and evaluation is a good tool for ensuring usability and likeability" of smart products. The challenge as Säde sees it is to find a combination of techniques or iteration cycles, which flow from one to another in an integrated way, sufficient to show the real benefits to management. He suggests that a good way to start is to "use well-focussed, easy to implement, low fidelity modelling techniques".

While this author would argue that the Ward (1994) study described earlier convincingly demonstrates the efficacy and value of low-fidelity prototypes for testing control/display layout designs, it falls short of evaluating the physical attributes of the interface such as tactile, auditory and some visual feedback. It is apparent that different types of prototype, and associated levels of fidelity, can provide different types of design information. The question now is which type of prototype to use and when.

6. Which type of prototype to use?

The most appropriate type of prototype to use depends on various factors. It is apparent that low-fidelity prototypes are good for testing the cognitive aspects of the design such as the layout of controls and displays (which embodies such design issues as sequence and importance of use of controls), but not physical aspects such as tactile, auditory and/or visual feedback. Testing such physical aspects requires a higher level of fidelity as provided by Types 3–7 of Table 1.

It can also be argued that if users are to appraise realistically the aesthetic attributes of products before they are actually produced, then it will require user appraisal of high-fidelity prototypes. This may conveniently be done in a second or third iteration of the design process cycle, at the same time as the evaluation of the appropriate usability attributes.

To illustrate the different strengths and costs of different levels of fidelity of prototypes, results from studies using *cardboard*, *touch screen* and *solid form* prototypes are given below for the following "products".

- Domestic lighting controller—cardboard mock-up (Ward, 1994).
- Domestic lighting controller—touch screen prototype (Kenney, 1994).
- time setting controls—touch screen prototype (Hall *et al*., 1999).
- EFTPOS terminal—solid form prototype.

6.1. DOMESTIC LIGHTING CONTROLLER

Kenney (1994) replicated Ward's 1991 study (Ward, 1994) to assess the value of using a higher fidelity (touch screen) prototype and to compare the findings with those of Ward's cardboard and foam mock-up of a domestic lighting controller. Figure 1 shows Ward's original mock-up with a sliding card with system responses which are visible through the rectangular "cut-out LCD" (liquid crystal display) above the two rows of "membrane"-type switches. Another sliding card toggles **RED** or **GREEN** "LEDs" (light-emitting diodes).

While Kenney's study involved several weeks of programming to create the two designs (the original design of Figure 1 and a revised design), the extra fidelity afforded by the touch screen implementation revealed more problems. A fuller description of both Ward's and Kenney's studies is given in Hall† (1997, 1999).

The high-fidelity prototypes revealed 29 (or 100%) more problems than the low-fidelity prototypes; their extra fidelity allowed instant visual feedback and more realistic interaction. A comparison of the average number of problems experienced (by each subject) for each prototype (low and high fidelity) and for each design (original and revised) is given in Table 2. Thus, it is evident that the higher-fidelity prototypes revealed more problems and gave greater feedback, but at the cost of the time and effort needed to develop the touch screen prototypes (see Section 7 for discussion).

6.2. TIME SETTING CONTROLS

Hall *et al.* (1999) describe the use of a PC touch screen to simulate various designs for setting the time on small consumer products like bedside clock radios, where the number of control buttons has to be kept to a minimum because of space constraints. In order to control for other design factors (such as layout, type of buttons, etc.) which differentiate such products on the market, a common interface was developed using the Visual Basic™ programming language where only the time-controlling buttons were different. There were four designs simulated.

1. Plus–Minus design with two buttons labelled: **+** and **−**.
2. Hour–Minute with two keys labelled: **Hours** and **Minutes**.
3. 10–1 with four buttons labelled: **10** and **1** for both Hours and Minutes.
4. Slow–Fast with two buttons labelled: **Slow** and **Fast**.

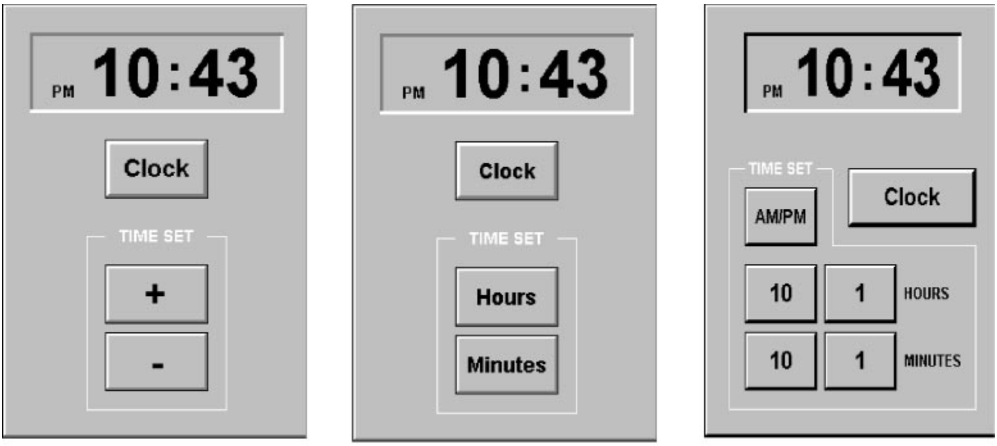
The first three designs are shown in Figure 2 to show the commonality in basic design but with different buttons. A fifth design, a telephone numeric keypad, was included as a control design based on the hypothesis that users are so familiar with telephones that it should be the most usable, i.e. the fastest with least errors.

Subjects had to perform five tasks which represented realistic situations where the time had to be reset because of say, a power black-out, or altered because of a change to, or back from, daylight saving time. The findings and design implications are fully discussed elsewhere (Hall *et al.*, 1999). Suffice it to say that the fidelity of this touch screen

TABLE 2
Average number (and SD) of problems revealed per user

Fidelity ↓	Design →	Original	Revised
Low—cardboard		8.5 (3.3)	2.5 (1.3)
High—touch screen		13.5 (2.9)	8.7 (4.9)

†The author supervised the studies of both Ward and Kenney.



(a) (b) (c)

FIGURE 2. (a) Plus–Minus. (b) Hours–Minutes. (c) 10–1.

prototyping revealed differences between the alternative designs for time setting on both objective (task time, number of button presses) and subjective (ratings) criteria. The telephone numeric keypad control design was, as expected, the fastest to use, easiest in recovering from an error, and rated highest in user satisfaction. However, in situations where space is a constraining factor the 10–1 design proved satisfying and quick to use, despite some initial difficulty in learning how to use it.

However, the resources required were reasonably high (computing expertise and time); but once one design had been developed on screen, the rest were easily developed at a marginal additional cost in resources. Further, with this type of prototyping it is possible to assess the optimum delay time and repetition rate when a button is held down to activate a fast forward function on some designs (e.g. Plus–Minus, Hours–Minutes, as is currently available on some domestic consumer products).

6.3. EFTPOS TERMINAL

There is a move in some supermarkets in Europe and the USA for customer scanning of groceries and self-payment at checkouts. The supermarket owners feel that the design of the EFTPOS (Electronic Funds Transfer, Point Of Sale) terminals should facilitate fast, error-free performance to reduce queues and possible frustration to customers.

Since such EFTPOS devices already exist and some of the control layouts are standardized, e.g. the numeric keypad, the emphasis on the use of prototypes may be more towards the physical design features such as location and length of swiping slots, the size and feel of the buttons, and the size and number of lines on the display. The relevant usability criteria become the time to complete the whole payment process (transaction time), the number of problems encountered during the task, and the customers' assessment (rating) of the ease and comfort of swiping their card.

FIGURE 3. (a) Prototype 1 (b) Prototype 3. (c) Prototype 5.

A study (unpublished) was recently conducted by the author using three existing and six new designs of EFTPOS terminals. Solid form prototypes were built with real keys and a four-line display which was controlled by a "checkout operator" using a five-position rotary switch. In response to the subject's key presses the switch was moved to the next position which illuminated each line of the display in turn. This was a quick and inexpensive mechanism which obviated the need for and cost of a programmable logic device built into the prototypes. Also, the real display of the three existing EFTPOS terminals was replaced with the above mechanism as an experimental control measure.

Three of the new, prototype terminals (1, 3 and 5) are shown in Figure 3. Three other prototypes (2, 4 and 6) were easily created by removing two given *panels* from each of those of Figure 3 and substituting two new *panels*. For example the horizontal swiping *panel* at the top of Prototype 1 was replaced by a short vertical one on the right-hand side to create Prototype 2, etc. In this way, there was some economy of scale in that only three prototype bodies and displays had to be formed.

The study revealed significant differences between the nine designs in terms of the average transaction time, and ratings of ease and comfort of use. Also, problems in using the prototypes were recorded because they highlight the design features which are related to longer transaction times and poorer ratings of design. Eighteen distinct problems were identified by 30 subjects drawn from a market research company database according to age, stature and handedness to be representative of the target user population. The problems were categorized into three mutually exclusive groups.

1. Hesitation or confusion showing cognitive difficulty (5).
2. Difficulty in action showing motor difficulty (5).
3. Errors or incorrect actions (8).

As expected, those designs for which there were a large number of problems also had a significantly long average transaction time ($r = 0.90$).

In this particular case the investment in producing the solid form prototypes was high, i.e. the cost of producing six prototypes and testing the nine designs was high. But the user testing of the prototypes revealed design information which clearly indicated which design features the sample of representative customers clearly liked, and for which they performed faster and with minimum errors. It was this performance information which would be required for this type of product. Subjective ratings of comfort and preference were also useful to complement the performance data.

7. Choosing the right prototype

7.1. RESOURCE COSTS AND SEVERITY OF PROBLEMS REVEALED

From Section 6.1, Table 2, it can be seen that the higher-fidelity prototypes revealed more problems which gave greater feedback about the problems in the original design, but at the higher resource cost of time to create the touch screen interfaces. Thus, there may well be a trade-off between the design information gained and the resources expended to gain that information.

However, it is not clear if the extra information gained with a higher-fidelity prototype is in proportion to the extra resources needed—not all problems are of equal magnitude or severity, and hence "cost". Some problems have more impact than others and/or are experienced by more of the users. Others may only occur with first use of the product, and once learnt may no longer be a problem, e.g. use of the American date format (month/day/year) in a country which uses the day/month/year format. (In this author's view there is no real justification not to use the appropriate date format for local markets, it often just requires a small software change.)

Nielsen and Mack (1994) when discussing the heuristic evaluation of "a prototype user interface for a system for internal telephone company use", the integrating system (IS), state that the severity of a usability problem is a combination of three factors: the *frequency* of the problem, the *impact* of the problem, and the *persistence* of the problem. While their "expert" (heuristic) evaluators were able to rate the severity of a usability problem on a five-point scale, sole reliance on such subjective rating may not always be warranted. Not only did user testing of the same IS prototype reveal four additional problems to the original 40 found through heuristic evaluation, but also it failed to reveal 23 (or 57%) of the original 40 found through heuristic evaluation.

Nielsen and Mack (1994) recommend that heuristic evaluation be used first to "clean up the interface and remove as many "obvious" usability problems as possible" and then perform user testing on the revised design. However, they do not say how many evaluators should be used in this iterative design process. They do conclude that "severity ratings from a single evaluator are too unreliable to be trusted … [but that] … ratings from three or four evaluators would seem to be satisfactory for many practical purposes".

When employing user testing to evaluate prototypes, it is very easy, and objective, to count the number of problems experienced by subjects. Also, in setting usability performance requirements it is usual to specify the percentage of users for whom performance is to be satisfactory. This value is usually 90 or 95% of the sample of users (subjects) involved. Therefore, in comparing the results of Ward (1994) and Kenney (1994), this author has defined a problem to be major if more than two subjects experienced the problem. This means that at least 80% subjects "performed" satisfactorily.

The problems for each prototype of the original design were reanalysed to see which were major according to the above definition. Of the total 54 problems found for the original design for both fidelities, only 21 were major. It can be seen in Table 3 that of all the major problems the low-fidelity prototype found 67% while the high-fidelity one found 90%.

However, of the total number of major problems revealed, only 15 *could* have been found with the low-fidelity prototype because of its reduced fidelity. Of those possible 15

TABLE 3
Number (& %) of all and possible major problems for the original design

Fidelity↓	Major →	All (21)	Possible (15)
Low—cardboard		14 (67%)	14 (93%)
High—touch screen		19 (90%)	15 (100%)

Note: A major problem is defined as a problem experienced by more than 20% of the subjects.

major problems the low-fidelity prototype found 14 or 93%, indicating the high efficacy of the low-fidelity prototype. The high-fidelity prototype did find one extra problem that the low-fidelity prototype could have possibly found but at the same time the low-fidelity prototype found two major problems that the high-fidelity prototype did not find (it only found 19 of 21, meaning that only 12 were found to be in common).

While this author acknowledges the role and importance of heuristic evaluation in an iterative design process, it would be interesting to conjecture if heuristic evaluation of Ward's original design would have revealed any differences between the two fidelities. In theory, it should not because the same design would be evaluated, albeit at low and high fidelities. However, there may be some form of interaction between the fidelity of design and the evaluators—this poses an interesting research question.

7.2. FACTORS TO BE CONSIDERED

From the earlier discussion and the results of the evaluation studies, using prototypes with different levels of fidelity and appropriate usability criteria, the type of prototype to be used in any particular design context will depend on a number of factors such as the following.

1. What product is being designed?
   If the product is a single function product with a simple interface like the domestic lighting controller it may be adequately tested using a low-level prototype such as a cardboard mock-up. If it is a multifunction product with a more complex interface it may require testing of several different types of prototypes. An approach like Nielsen's horizontal and vertical prototyping (Nielsen, 1993) could be utilized to examine design issues in this case.
2. What design information is required?
   If the information required is about the optimum layout of controls, then several low-fidelity mock-ups may be sufficient. If the information required is about the optimum feel (tactile feedback, etc.) of keys/buttons/switches, then a higher level of fidelity is required and preferably such testing should follow that for determining the optimum layout.
3. What stage of the design process?
   Generally, the later the stage in the design process the more appropriate it is to have higher-fidelity prototypes. However, there may be some situations where it is still appropriate to test low-fidelity mock-ups late in the design process.

4. What resources are available?
   It should be within the resources of most design companies to build low-level prototypes such as paper or cardboard mock-ups. If the company is larger it may have staff with computing expertise and therefore the creation of touch screen prototypes may become cost-effective. The same may be said of model makers on staff and the development of solid form prototypes.
5. What are the likely costs of developing a poor design?
   In the case of products for which there are large production costs and/or large-scale deployment, like EFTPOS and similar products, the business case to produce higher-fidelity prototypes may be more easily made. However, the author would argue for first testing lower-fidelity prototypes to deduce the optimum design from a cognitive ergonomics point of view before proceeding to test the physical ergonomics aspects of the design.

8. A design strategy

There will not be a single design strategy which will suit all products, design contexts or budgets. It has been argued earlier that to achieve a good or usable design, the design process should be both user-centred and iterative. This paper has focussed particularly on user testing as a form of evaluation of designs, but Nielsen and Mack (1994) have strongly argued for a combination of heuristic evaluation and user testing of prototypes, with which this author agrees. However, it is extremely difficult to construct, let alone describe, a unified design strategy to cover all design situations. Therefore, this author suggests that an appropriate design strategy based on user testing could be as follows.

1. Select the lowest level of fidelity deemed appropriate for the product being designed —see above.
2. Select a small representative sample of target users as subjects. There are previous studies (Virzi, 1990; Nielsen, 1990; Ward, 1994) which indicate that only a small number of subjects (around five to seven) are required to find the majority (around 80%) of the problems. The dotted line in Figure 4, taken from Ward (1994), shows the cumulative percentage of problems revealed by subjects in the actual order they participated in the study. The solid line indicates the theoretical, average rate of revealing problems (with the standard deviation shown by the short vertical bars) based on a reanalysis of the data. Hall (1997) suggests that the cumulative number of new problems found with each subject be plotted against the number of subjects and that, as the curve starts to flatten off, the testing of the prototype be stopped and the design revised, based on the design information feedback, and then tested again (cf. Point 6 below).
3. Select representative tasks to perform (cf. ISO 1998).
4. Select appropriate usability criteria to measure—based on “problems” or formal ISO (1998) criteria.
5. Run user trials in a scientific manner—control for biases in test procedure, stress to the subject (user) that the product is being tested not them, and use think aloud technique with post-test interview of subject.

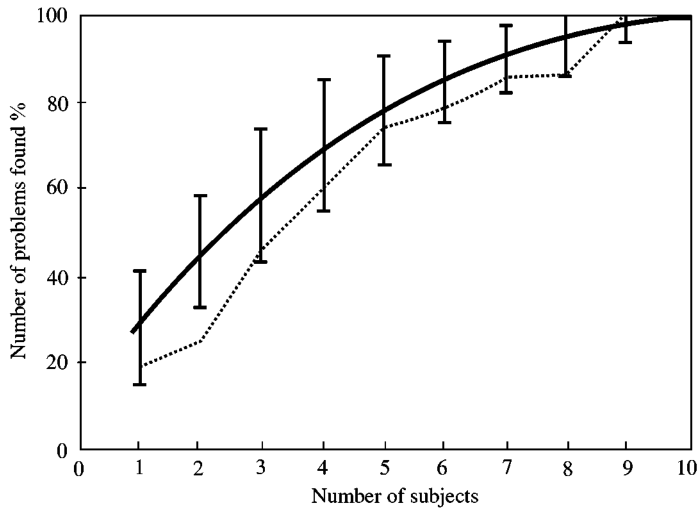


FIGURE 4. Relationship between number of subjects used and the cumulative number (%) of problems found (after Ward, 1994).

6. Redesign (based on problems found) with higher level of fidelity.
7. Repeat process as necessary.

While this strategy has been based on user testing as the evaluation tool, heuristic evaluations may prove more cost-effective than user testing at certain, appropriate stages or with certain types of prototypes. Further research may need to be done to determine when and under what circumstances heuristic evaluation and/or user testing are best suited. However, a human–computer interaction or usability expert, with a good understanding of the design "problem", should be able to provide appropriate guidance on evaluation during the design process.

9. Conclusions

Based on the studies of the author and others cited above, the following can be reasonably concluded.

1. It is possible to achieve good design information from user testing of low fidelity prototypes early in the design process.
2. The fidelity of the prototype determines the type of information that can be gained from user testing.
3. The value of design information depends on the fidelity of the prototype tested.
4. There is a need for cognitive design information to be gathered and dealt with before physical issues such as tactile feedback.
5. The economic benefit of the prototyping depends on the resources deployed for gathering the design information relative to the potential cost of producing a poor design.

References

Adams, A. & Hall, R. R. (1992). Design of visual displays in ticket vending machines. *Work With Display Units, Abstract Book of the Third International Scientific Conference on Work with Display Units*, pp. E24–25, Berlin, 1–4 September.

Baecker, R. M. & Buxton, W. A. S. (1987). *Readings in Human–Computer Interaction: A Multidisciplinary Approach*. San Mateo CA: Morgan Kaufmann.

Blatt, L. A. & Knutson, J. F. (1994). Interface guidance systems. In J. Nielsen & R. L. Mack, Eds. *Usability Inspection Methods*, pp. 351–384. New York: John Wiley.

Cushman, W. H. & Rosenberg, D. J. (1991). *Human Factors in Product Design*. Amsterdam: Elsevier.

De Vries, G. & Johnson, G. L. (1999). Spoken help for a car stereo: an exploratory study. In P. Jordan & W. S. Green, Eds. *Human Factors in Product Design*, pp. 124–137. London: Taylor & Francis.

Dix, A. Finlay, J. Abowd, G. & Beale, R. (1993). *Human–Computer Interaction*. London: Prentice Hall.

Eason, K. D. (1992). *The development of a user-centred design process: a case study in multidisciplinary research*. Inaugural lecture, 14 October, 1992, Loughborough UK: Loughborough University of Technology.

Evans, M. & Moore, C. (1991). *Technology: blessing or curse?* National Social Science Survey Report, 2(5), pp 1–4. Canberra: Research School of Social Sciences ANU.

Gould, J. D. (1995). How to design usable systems. In R. M. Baecker, J. Grudin, W. A. S. Buxton & G. S. Greenberg, Eds. *Readings in Human–Computer Interaction*, 2nd edn, pp. 93–121. San Francisco: Morgan Kaufmann Publishers.

Gould, J. D. & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, **28,** 300–311.

Hall, R. R. (1997). Ergonomics, design and new technology. The Cumming Memorial Lecture. *Productivity, Ergonomics and Safety, Proceedings of the 33rd Annual conference of the Ergonomics Society of Australia*, pp. 53–62, Gold Coast, Queensland, November. pp. 53–62. Canberra: Ergonomics Society of Australia Inc. & at *www. detir.qld.gov.au/hs/ergo97/hall2.pdf*.

Hall, R. R. (1999). Usability and product design: a case study, In P. Jordan & W. S. Green, Eds. *Human Factors in Product Design*, pp. 85–91. London: Taylor & Francis.

Hall, R. R. Zinser, S. & Keller, P. (1999). The usability of time setting functions on small electronic consumer products—a test. *International Journal of Cognitive Ergonomics*, **3,** 101–114.

ISO (1998). *International Standard 9241-11: Ergonomic requirements for office work with visual display terminal (VDTs)—Part 11: Guidance on Usability*. Geneva: International Organisation for Standards.

ISO (1999). *International Standard 13407: Human-centred design processes for interactive systems*. Geneva: International Organisation for Standards.

Jordan, P. (1998). *An Introduction to Usability*. London: Taylor & Francis.

Kahmann, R. & Henze, L. (1999). Usability testing under time-pressure in design practice. In P. Jordan & W. S. Green, Eds. *Human Factors in Product Design*, pp. 113–123. London: Taylor & Francis.

Karwowski, W. Chase, B. Gaddie, P. Lee, W. & Jang, R. (1997). Virtual reality in human factors research and human factors of virtual reality. In P. Seppala, T. Luopajarvi, C. H. Nygard & M. Mattila, Eds. *From Experience to Innovation, Proceedings of the 13th Triennial Congress of the International Ergonomics Association, Tampere, Finland, Vol.* 2, pp. 53–55, 29 June–4 July. Helsinki: Finnish Institute of Occupational Health.

Kenney, T. (1994). *A comparison between different interface evaluation methodologies: a test*. Bachelor of Engineering thesis, School of Computer Science and Engineering, University of New South Wales, Sydney, Australia.

Maguire, M. C. Butters, L. M. & McKnight, C. (1994). Usability issues for buyers and users of home electronic products. In R. Oppermann, S. Bagnara & D. Benyon, Eds. *Human–Computer Interaction: From Individuals to Groups in Work, Leisure, and Everyday Life, Proceedings of ECCE 7, 7th European Conference on Cognitive Ergonomics*, pp. 117–133, Bonn, 5–8 *September*. Sankt Augustin: Gesellschaft fur Mathematik und Datenverarbeitung mbH.

MEISTER, D. (1990) System effectiveness testing. In G. SALVENDY, Ed. *Handbook of Human Factors*, pp. 1271–1297. New York: John Wiley.

METZ, S. RICHARDSON, R. & NASIRUDDIN, M. (1987). RAPID—software for prototyping user interfaces. In: *Proceedings of the Human Factors Society 31st Annual Meeting*, pp. 1000–1004. Santa Monica CA: Human Factors Society.

NIELSEN, J. (1990). Big paybacks from "discount" usability engineering. *IEEE Software*, **7,** 107–108.

NIELSEN, J. (1993). *Usability Engineering*. London: Academic Press.

NIELSEN, J. & MACK, R. L. (1994). *Usability Inspection Methods*. New York: John Wiley.

NORMAN, D. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

NORMAN, D. (1998). *The Invisible Computer*. Cambridge, MA: MIT Press.

NORMAN, D. & DRAPER, S. W. (1986). *User-Centered System Design*. Hillsdale NJ: Lawrence Erlbaum.

PORTER, S. & PORTER, J. M. (1999). Designing for usability; input of ergonomics information at an appropriate point, and appropriate form, in the design process, In P. JORDAN & W. S. GREEN, Eds. *Human Factors in Product Design*, pp. 15–25. London: Taylor & Francis.

PREECE, J. ROGERS, Y. SHARP, H. BENYON, D. HOLLAND, S. & CAREY, T. (1994). *Human–Computer Interaction*. Wokingham, UK: Addison-Wesley.

ROODEN, M. J. (1999). Prototypes on trial. In: P. JORDAN & W. S. GREEN, Eds. *Human Factors in Product Design*, pp. 138–150. London: Taylor & Francis.

SÄDE, S. (1999). Representations of smart product concepts in user interface design. In P. JORDAN & W. S. GREEN, Eds. *Human Factors in Product Design*, pp. 64–72. London: Taylor & Francis.

SEMINARA, J. L. & GERRIE, J. K. (1966). Effective mockup utilisation by the industrial design-human factors team. *Human Factors*, **8**, 347–359.

SHACKEL, B. (1959). A note on panel layout for numbers of identical items. *Ergonomics*, **2**, 247–253.

SHACKEL, B. (1986). Ergonomics in design for usability. In M. D. HARRISON and A. F. MONK, Eds. *People & Computers: Designing for Usability*, pp. 44–64. Cambridge: Cambridge University Press.

SHNEIDERMAN, B. (1998). *Designing the User Interface: Strategies for Effective Human–Computer Interaction, 3rd edn*. Reading MA: Addison Wesley.

STANTON, N. A. & YOUNG, M. S. (1999). *A Guide to Methodology in Ergonomics*. London: Taylor & Francis.

THIMBLEBY, H. (1991). Can anyone work the video? *New Scientist*, **23**, 40–43.

VIRZI, R. A. (1990). Streamlining the design process: running fewer subjects. *Proceedings of the Human Factors Society 34th Annual Meeting*, pp. 291–294. Santa Monica, CA: Human Factors Society.

VIRZI, R. A. PENN, D. TULLIS, T. S. & GREENE, S. L. (1990). The uses of prototyping in user interface design and evaluation. *Proceedings of the Human Factors Society 34th Annual Meeting*, pp. 264–266. Santa Monica, CA: Human Factors Society.

WARD, S. (1994), Getting feedback from users early in the design process: a case study. In N. ADAMS, N. COLEMAN & M. STEVENSON, Eds. *Ergonomics: the Fundamental Design Science, Proceedings of the 30th Annual Conference of the Ergonomics Society of Australia*, pp. 22–29, 4–7 December. Canberra: The Ergonomics Society of Australia.

WRIGHT, P. & MONK, A. F. (1989). Evaluation for design. In H. SUTCLIFFE & K. L. MACAULAY, Eds. *People and Computers V, Proceedings of the 5th British Computer Society, Human–Computer Interaction Special Interest Group Conference*, pp. 345–358, 5–8 September Cambridge: Cambridge University Press.