What price ergonomics?

Neville A. Stanton[1] and Mark S. Young[1]

Abstract

Ergonomists have a say in the design of almost everything in the modern world, but there is little evidence that their methods actually work. Here is an evaluation of those methods and of the worth of ergonomics in design.

Introduction

Engineers make things that are useful to people. In collaboration with designers, ergonomists make things that are usable by people. The concept of usability means making artefacts easy, efficient and comfortable to use (anything from a corkscrew[1] to a control room in a nuclear power station[2]). Most people have experience of poorly designed objects. At best they cause frustration and annoyance (for example when a video recorder fails to record your favourite programme). At worst they can lead to injury or even death (as in the release of radioactive material from a nuclear reactor).

'User friendly', 'easy to use' and 'ergonomically designed' are commonly found claims for domestic appliances and devices, particularly in the advertising business. For instance, an advertisement for a well-known music system boasts that it is 'Technologically advanced, yet remarkably easy to use'. But only a few people — including some, but not all, of those behind such advertisements — know what is involved in an ergonomic assessment.

Ergonomists do more than simply design chairs (a common misconception). Rather they are involved in various aspects of the design and evaluation of a host of products. They work on the specification of user requirements, the development of design guidelines, the evaluation of device prototypes, and the implementation and analysis of user trials. Traditionally, ergonomics has had most influence on the physical characteristics of design. For example, in conceiving of a new radio-cassette machine, ergonomists might advise on the buttons: their size and placing, and the reach and force required to activate them. The ergonomist draws on a substantial body of research findings and hard data, such as human anthropometrics — tables of data that contain information about the range of human physical attributes and characteristics[3].

Nowadays, ergonomists also advise on interactive devices, such as so-called WIMP (Windows, Icons, Menus and Pointers) interfaces. Advice includes the form that the interface should take, relating the grouping of functions and device elements to the goals of the user, predicting user satisfaction, and estimating task performance time and the error potential of the device. These analyses are used to select from a choice of alternative designs and to improve the chosen option. But analyses of human cognitive functioning do not rest upon hard data to the same extent as anthropometrics[4]. Questioning these techniques is our purpose here.

Despite the proliferation of ergonomic methods in research[5], teaching[6] and industry[7], there is little substantive evidence that these methods actually work[8]. Part of the reason is that researchers tend to stick to one or two that they know, trust and, in many cases, have developed themselves. The main point here, however, is that methods for predicting the usability of devices have become so entrenched that their validity is simply assumed and seldom tested. These methods are used in good faith, in the belief that the results are reliable and valid. But that assumption is questionable, and we find that some approaches do not do all that is claimed for them.

Reliability and validity

The objective way to see whether ergonomic methods work is to assess their reliability and validity. Methods can be reliable without being valid, but they cannot be valid without being reliable. We can liken reliability and validity to the accuracy of a rifle marksman using a gunsight. Reliability refers to the grouping of the shots (the error of the marksman), whereas validity refers to the closeness of each shot to the centre of the target (the error of the gunsight). If ergonomic methods are shown to be adequately reliable and valid, they may be used with increased confidence.

In a series of studies, we took ten of the more popular ergonomic methods, and subjected them to tests of reliability and validity (details of the methods are found in ref. 1, and ergonomic textbooks). The ergonomic methods each predicted different aspects of usability, from objective performance measures (such as reaction time and errors) through to subjective criteria (such as user satisfaction). The methods evaluated included classical approaches (such as interviews and observation), as well as some of the more modern techniques (such as SHERPA, the Systematic Human Error Reduction and Prediction Approach). Only a few people were involved in the study: eight were trained in each of the methods and 30 were observed using a radio-cassette machine, the device we chose to use. But these numbers are realistic in terms of the kind of studies carried out by ergonomists. Our analyses produced between 500 and 1,650 data points for each method, depending upon the method used, which were classified as hits, misses, false alarms and correct rejections.

The results, which appear in Table 1, show huge variability in both reliability and validity of the various methods. The validity statistics were calculated using either Pearson or Phi (a correlation coefficient for dichotomous data). For the latter PhiMax is also given, as this shows the maximum value that could be achieved, given the marginal distributions[9]. Interestingly, two of the methods return small but negative correlations, meaning that the predicted behaviour was the opposite of that observed.

Table 1: Reliability and validity of ergonomic methods

Ergonomic Method	Reliability	Validity	PhiMax
Checklists	0.307	0.206*	0.659
Heuristics	0.471	0.087*	0.464
Hierarchical task analysis	0.226	0.188*	0.937
Interviews	0.449	-0.112*	0.422
Keystroke level model	0.916	0.769*	N/A
Link analysis	0.830	0.356*	0.572
Observation – errors	0.890	-0.141*	0.343
Observation – time	0.890	0.729*	N/A
Questionnaires	0.578	0.615*	N/A
Repertory grids	0.562	0.078	0.681
SHERPA	0.392	0.238*	0.922

Reliability was calculated by Pearson's product moment correlation coefficient. Validity was calculated by Phi (those with associated PhiMax values) and Pearson's product moment correlation coefficient. Asterisk indicates that validity value was statistically significant.

So is it worth using these methods at all? Thus far in the argument, the answer has to remain 'perhaps'. The reason is that ergonomists not only have an agenda of making life in our technological world easier, safer and more enjoyable, but also that of helping to maximize manufacturer revenue and minimize costs. Designing a product in accordance with usability principles can increase profits; and, given the sums of money involved in product development, even methods with relatively low validity could offer returns.

Methods that can predict good or bad designs should save time and money in product development. They might increase sales and profits, and reduce need for training and after-sales support. But those benefits have to be set against the costs of setting up an ergonomics laboratory, conducting the methods and evaluating iterative prototypes. So does the relative inaccuracy of the methods reduce these benefits beyond the point where they are worth the effort and cost? That is, does it cost more to use a method than is gained from the end result?

Cost-benefit analysis

Claims as to the cost-effectiveness of ergonomic methods normally centre on the results of cost-benefit analysis. Here the approach is to calculate the cost of applying the method (in terms of person-hours, materials and so on) and subtract this from the estimated savings generated by the improvement in design. The net figure is proposed as the benefit brought about by using ergonomic methods.

The figures can look good. According to one textbook example[10], the cost of a usability assessment of the interface for new computer software (US$171,445) compares very well to the potential savings of the new design ($592,635) in terms of

the reduced costs of training and after-sales support in the first year. On the face of it, then, in this instance ergonomic methods provide a profit of more than $400,000.

The estimate of likely economic benefits does not however attempt to take the accuracy of the ergonomic methods into account. The application of ergonomic methods (if inaccurate) could fail to predict true problems with device design or, worse still, identify problems that do not exist. We argue that the degree to which the financial benefits will be realized will depend upon the accuracy of the methods used. Using the reliability and validity data we acquired, the net benefit of each method can be calculated from its utility U (that is, its practical usefulness couched in financial terms): U = (validity $\times$ prospective increase in value) - costs. A recalculation of the textbook cost-benefit analysis would find a substantial reduction in the benefit by including the validity data. It is therefore essential that these data are considered when addressing utility, as they show how much variability can be accounted for in ergonomic methods.

The practical context here is that, in product design, a manufacturer is typically faced with the prospect of choosing between ten or more designs. Given this degree of choice, how are they to arrive at the best decision? They could flip a coin or rely upon intuition. Alternatively, they could use the methods to predict differences in user performance to select the best design. The difference between a good and poor design could be worth millions of dollars to a manufacturer over the lifetime of a product.

For the purpose of testing the formula with the computer software example[10], we can estimate that choosing a good design over an average design will increase profitability by upwards of $500,000 per year (a rounding down of the over-precise figure provided by the example). Similarly, we use $100,000 as the cost of making prototypes and assessing the device. The same costs are used in both analyses to show the effect of validity on utility. In our analysis, the method with the lowest validity value was use of repertory grids; that with the highest was the keystroke level model (KLM). The contrast of utilities brought about by these two methods, with respective validity values of 0.078 and 0.769, can be estimated as: U (repertory grids) = (0.078 $\times$ $500,000) - $100,000 = $61,000; U (KLM) = (0.769 $\times$ $500,000) - $100,000 = $284,500. So some methods may return negative values in the first year, whereas others show a healthy return on investment.

A range of factors can affect the selection of ergonomic methods to be applied in evaluating specific products, including time and resources available, efficacy of the methods, criteria by which the product is to be evaluated, and access to a 'test' sample of end-users. Utility analysis can help make some of these judgements more explicit and more objective. It can help in the selection of methods, show where effort may be best expended, and make the design process more effective. Experience from the field of personnel selection shows that for utility analysis to be effective it must be easy to use and the derivation must be transparent — people using the formula must be able to see how it works and carry out the calculations without having a PhD in mathematics.

We can provide the reliability and validity data, the costs of conducting each method, and the appropriate procedure for conducting a cost-benefit analysis. Product developers will have their own budgets for sales and training, and these will differ

depending on whether the product is a vending machine or a video recorder. So it is up to the product development teams concerned to make their own calculations and decide whether ergonomic methods are actually worth using.

Our aim here has not been to condemn ergonomic methods or ergonomics as a discipline. Rather, our take-home message is that, if ergonomists are really to improve the design of artefacts, their methods have to be shown to be both useful and usable.

Acknowledgements.

References

1. Stanton, N. A. *Human Factors in Consumer Products* (Taylor & Francis, London, 1998).
2. Stanton, N. A. *Human Factors in Nuclear Safety* (Taylor & Francis, London, 1996).
3. Pheasant, S. T. *Bodyspace — Anthropometry, Ergonomics and Design* (Taylor & Francis, London, 1986).
4. Dowell, J. & Long, J. T. *Ergonomics* **41**, 126–139 (1998).
5. Salvendy, G. *Handbook of Human Factors and Ergonomics*, 2nd edn (Wiley, New York, 1997).
6. Wilson, J. & Corlett, N. *Evaluation of Human Work*, 2nd edn (Taylor & Francis, London, 1995).
7. Jordan, P. W., Thomas, B., Weerdmeester, B. A. & McClelland, I. L. *Usability Evaluation in Industry* (Taylor & Francis, London, 1996).
8. Stanton, N. A. & Young, M. *Appl. Ergonomics* **29**, 41–54 (1998).
9. Guilford, J. P. *Psychometric Methods* (McGraw-Hill, New York, 1957).
10. Bias, R. G. & Meyhew, D. J. *Cost-Justifying Usability* (Academic, Boston, 1994).