

Is Utility in the Mind of the Beholder? A Study of Ergonomics Methods

Neville Stanton & Mark Young
Department of Psychology
University of Southampton
Highfield
Southampton
SO17 1BJ
UK

This paper reviews the use of ergonomics methods in the context of usability of consumer products. A review of the literature indicated that there is upward of 60 methods available to the ergonomist. The results of the survey indicated that questionnaires, interviews and observation are the most frequently reported methods used. Ease of use of the methods was dependent upon type of method used, presence of software support and type of training received. Strong links were found between questionnaires and interviews as a combined approach as well as with HTA and observation. However, a questionnaire survey of professional ergonomists found that none of the respondents had any documented evidence of the reliability and validity of the methods they were using. A study of training people to use ergonomics' methods indicated the different requirements of the approaches, in terms of training time, application time and subjective preferences. An important goal for future research is to establish the reliability and validity of ergonomics methods.

KEYWORDS: Methods, techniques, training, preferences.

Ergonomics Methods

There appears to be a growing number of texts in recent years describing, illustrating and espousing a plethora of ergonomics methods (Diaper, 1989; Kirwan & Ainsworth, 1992; Kirwan, 1994, Corlett & Clarke, 1995; Wilson & Corlett, 1995; Jordan, Thomas, Weedmeester & McClelland, 1996). The rise in the number of texts reporting on ergonomics methods may be seen as a response to the requirement for more inventive approaches for assessing the user and their requirements. In many ways this may be taken to mean that the call for user-centred design has been taken seriously by designers. However, this success has forced the ergonomics community to develop methods to assist the design of products and devices. This demand seems to have resulted in the pragmatic development of methods having priority over scientific rigour. In a recent review of ergonomics methods, Stanton & Young (1995) identified over 60 methods available to the ergonomist. The abundance of methods might be confusing for the ergonomist, Wilson (1995) goes as far as to suggest that a

"...method which to one researcher or practitioner is an invaluable aid to all their work may to another be vague or insubstantial in concept, difficult to use and variable in its outcome." (p. 21)

This quote highlights the fact that most methods are used by their inventors only. An ergonomist approaching a collection of methods might ask questions such as:

1. Which method is appropriate?
2. How long does it take to train people to use the method?
3. How long do people take in applying the method to evaluate a device?
4. What are the relative benefits of one method over other methods?

Despite the proliferation of methods, there are few clues in the literature available for ergonomists to enable them to answer these questions. The purpose of this paper is to attempt to address some of these issues and focus our attention on usability measures and has focused largely upon methods that examine user activity and behaviour. Some researchers have attempted to address the first question. For example, an overview of the methods is presented in most of the books (Diaper, 1989; Kirwan & Ainsworth, 1992; Kirwan, 1994, Corlett & Clarke, 1995; Wilson & Corlett, 1995; Jordan, Thomas, Weedmeester & McClelland, 1996). Stanton and Baber (1996) reduce the selection of methods down to four basic factors:

- the stage of the design process

- the form that the product takes
- access to end users
- pressure of time.

One or more of these factors will determine which methods are appropriate. From this analysis, Stanton & Baber (1996) argue that it is not surprising that the checklist is the ubiquitous ergonomics method, as it is the only method that is independent of these factors. We have analysed the methods reported in the six texts cited above in order to determine how many methods are cited, and the emphasis is given to generality of the domain of application, as well as looking for evidence to help us answer the four questions we raise regarding ergonomics methods. This analysis, albeit subjective, is presented in table 1.

TABLE ONE ABOUT HERE

Table 1. An overview of six ergonomics' methods books.

As table 1 shows, all but one of these texts are multi-authored and all but one were produced in the last four years, although two are second editions. The number of methods contained within the texts ranges from six to 48. Most of the texts are general in nature. Three of the texts contain validation studies, but these are sparse and only apply to a few of the methods mentioned. Finally, none of the texts contain any description of studies that relate to the acquisition of the method or, apart from Kirwan (1994), the relative merits of one method over another. In order to answer the four questions raised earlier, this paper is organised into two sections: a survey and review of ergonomics methods and a study of training in ergonomics methods. The questions are addressed in a general discussion at the end of the paper.

Survey of ergonomics methods

In order to evaluate ergonomists' practices in the use of methods, it was necessary to conduct a survey of professional ergonomists. In a recent survey of 6 organisations, conducted by Baber & Mirza (1995), a very limited range of methods were reportedly used in product evaluation. The methods typically used consisted of questionnaires, interviews, observations, checklists and heuristics. Baber & Mirza (1995) report that the frequency with which a method is used is highly dependent upon its ease of use, and most respondents reported combining 3 or 4 methods to obtain an overall picture of the product under evaluation. We decided to follow up this survey with our own to

question respondents' experience with ergonomics methods on the following dimensions:

- the type of techniques used
- relevant instructional documentation
- evidence of validity
- evidence of reliability
- the costs and benefits associated with the use of the technique
- the usability of the technique
- the potential applicability of the technique
- overall evaluation and conclusions.

The full survey instrument is contained within Stanton & Young (1996) and is available upon request (see figure 1).

Participants

Eighteen self selected respondents from a pool of 163 members of The Ergonomics Society listed on the professional register, were asked to comment on their experience with ergonomics methods. Participants were predominantly male (five were female) and working in a commercial environment (only two academics responded). Reported age range was 28-59 years, and only three responded from outside the UK (two from the USA and one from the Netherlands). Four respondents were educated to Bachelors level, nine to Masters level and five to Doctorate.

Design

The questionnaire was developed according to the principles of design proposed by Youngman (1982). The main stages involved in the design of the questionnaire were as follows:

- brainstorming
- review of literature
- drafting items
- pilot questionnaire
- review questionnaire with pilot respondents
- restructure questionnaire
- conduct survey

Procedure

Respondents were introduced to the aims of survey in an introductory paragraph as follows: "We are particularly interested in which methodologies are used, what they may be applied to, and whether such techniques are useful". First they were asked to complete some biographical details. Next they were required to consider a list of ergonomics methods, and only asked to complete information on the ones that they had experience with. For each method, respondents were asked to complete details on the utility of the method (i.e. the frequency of use, the ease of use, the cost-benefit ratio, the time taken and the availability of software support), their background experience with the method (i.e. the nature of training received, number of years used, the application of the method and links with other methods) and any reported evidence about the method (i.e. relationship to standards, reports of reliability, reports of validity and original source of information on method). Finally, respondents were asked to consider if there were any additional methodologies that were not covered in the original list and to complete extra response sheets if applicable. Respondents were thanked for their time and asked to indicate if they wanted a copy of the results. The results were sent to all respondents who requested them, i.e. all but one of the respondents.

Materials

The questionnaire was paper-based and respondents were required to enter their responses into the columns as shown in figure 1.

FIGURE ONE ABOUT HERE

Figure 1. An example of the questionnaire response sheet

Analysis

Only 18 completed forms were returned, which is rather disappointing, even for a postal survey. However, the results should provide a rough indication of some ergonomists' perception of the methods under question. The data were analysed on SPSS using the Kruskal-Wallis 1-way ANOVA followed by the Mann-Whitney U-test.

Results and Discussion of Survey

Details of 27 methods used by ergonomists are presented in table 2. Of the 27 reported methods, only 11 were reported by 6 or more respondents (i.e. from Mock-ups to Walkthroughs in table 2) and only these methods were treated to statistical analysis. In general terms, the results of the survey seem to confirm the analyses undertaken thus far. First, there were no references by the respondents to reported evidence for reliability or validity in the literature which concords with our earlier investigations

(Stanton & Young, 1995). Second, the respondents' evaluations of the techniques is consistent with our own experience.

TABLE TWO ABOUT HERE

Table 2. Summary of survey responses (variables "Frequency" through "Time" are medians of respondents' ratings; "Software" and "Training" variables are frequency counts; "Years" is mean reported use)

The statistical analysis of the respondents' ratings of the techniques revealed three interesting and statistically significant results. First, some methods were rated as easier to use than others (Chi-square, corrected for ties, = 33.0595; $p < 0.0005$). Checklists were rated as significantly easier to use than simulation (Z , corrected for ties, = -3.3994; $p < 0.001$), guidelines were rated as significantly easier to use than prototyping (Z , corrected for ties, = -2,578; $p < 0.01$) and interviews were rated as significantly easier to use than mock-ups (Z , corrected for ties, = -2.1381; $p < 0.05$). We noted earlier that only a limited range of methods are used but there is no guarantee that these are the most appropriate. Baber & Mirza (1995) report that product designers tend to restrict their methods to interviews, observation and checklists (which was confirmed in our study). Similar to the report by Baber & Mirza (1995), our finding suggests that this is likely to be due to the ease of applying the methods. Second, the reported ease with which methods are applied depends upon whether software support is used. The results show that where no software support is used, the method is rated as easier to use than where software support is provided (Z , corrected for ties, = -2.6597; $p < 0.01$). Although perhaps this is a counter-intuitive finding, our own experience suggests that software can make even a relatively easy method quite complex and cumbersome. With some irony we would suggest that developers of ergonomics software cannot afford to ignore ergonomics in the design of their product! However, it is our experience that software can make the ergonomist's activities more efficient in the long term. Finally, the data suggest that users of ergonomics methods perceive differences in ease of use of methods depending upon the level of training they have received (Chi-square, corrected for ties, = 6.0639; $p < 0.05$). Those who have received no training rate the methods as easier to use than those who have received informal training (Z , corrected for ties, = -1.9919; $p < 0.05$). We suggest that this result is probably due some misconception regarding how to use the method and would recommend formal training in any approach used.

Table 3 shows the reported links of each method with other methods. As can be seen there are strong links between questionnaires and interviews as a combined approach.

Interviews are also linked with observation, HTA and prototyping. Observation is linked with questionnaires and HTA. Finally, HTA is linked with HEART. Some of these links with other methods are by necessity. For example, HTA requires observation and interviews as necessary prerequisites to collect and verify the analysis. Similarly, HEART requires HTA before it can be conducted. Other links between methods are shown in lighter shading on table 3.

TABLE THREE ABOUT HERE

Table 3. Reported links between methods.

Table 4 shows the uses to which the methods are put. The dark shading shows that questionnaires, interviews and observation are the principal methods used in data collection. The medium shading shows that simulators, computer simulation, interviews and repertory grids are used largely for design activities. It also shows that simulators are used for assessment activities and that repertory grids are used in validation activities. The light shading indicated occasional usage of other methods for a range of activities with no clear pattern.

TABLE FOUR ABOUT HERE

Table 4. Reported applications of methods.

A brief review of some popular methods

Analysis of the content of the six texts cited in table 1 and a review of the responses on the survey led us to select 11 methods for further consideration. These methods were selected based upon our analysis that these are a representative spread of methods that are currently being used to evaluate human-machine performance and assess the demands and effects upon people (Wilson, 1995). Methods selected for analysis were as follows :

- Heuristics
- Checklists
- Observation
- Interviews
- Questionnaires
- Link analysis
- Layout analysis
- Hierarchical Task Analysis
- Predictive Human Error Analysis
- Repertory grids
- Keystroke Level Model

The aim of the review was to evaluate a device with each method in turn to determine the inputs (those materials and activities required to start the evaluation), processes (the activities of evaluation) and outputs (the resultant information supplied by the evaluation). A brief review of each of the methods is given in the following section. The review provides the basis for considering methods in a little more detail before the more formal evaluations.

Heuristics (Nielsen, 1992)

Heuristics require the analyst to use their judgement, intuition and experience to guide them on product evaluation. This method is wholly subjective and the output is likely to be extremely variable. In favour of the heuristic approach is the ease and speed with which it may be applied. Several techniques incorporate the heuristic approach (e.g. checklists, guidelines, PHEA) but serve to structure heuristic judgement.

Checklists (Ravden & Johnson, 1989; Woodson, Tillman & Tillman, 1992)

Checklists and guidelines would seem to be a useful aide memoire, to make sure that the full range of ergonomics issues have been considered. However, the approach may suffer from a problem of situational sensitivity, i.e. the discrimination of an appropriate item from a non-appropriate item largely depends upon the expertise of the analyst. Nevertheless, checklists offer a quick and relatively easy method for device evaluation.

Observation (Drury, 1990; Kirwan & Ainsworth, 1992; Baber & Stanton, 1996a)

Observation is perhaps the most obvious way of collecting information about a person's interaction with a device; watching and recording the interaction will undoubtedly inform the analyst of what occurred on the occasion observed. Observation is also a deceptively simple method, one simply watches, participates in, or records the interaction. However, the quality of the observation will largely depend upon the method of recording and analysing the data. There are concerns about the intrusiveness of observation, the amount of effort required in analysing the data and the comprehensiveness of the observational method.

Interviews (Cook, 1988; Sinclair, 1990; Kirwan & Ainsworth, 1992)

Like observation, the interview has a high degree of ecological validity associated with it: if you want to find out what a person thinks of a device you simply ask them. Interviewing has many forms, ranging from highly unstructured (free-form discussion) through focused (a situational interview), to highly structured (an oral questionnaire). For the purposes of device evaluation, a focused approach would seem most appropriate. The interview is good at addressing issues beyond direct interaction with

devices, such as the adequacy of manuals and other forms of support. The strengths of the interview are the flexibility and thoroughness it offers.

Questionnaires (Brooke, 1996)

There are few examples of standardised questionnaires appropriate for the evaluation of consumer products. However the Software Usability Scale (SUS) may, with some minor adaptation, be appropriate. SUS comprises 10 items which relate to the usability of the device. Originally conceived as a measure of software usability, it has some evidence of proven success. The distinct advantage of this approach is the ease which the measure may be applied. It takes less than a minute to complete the questionnaire and no training is required.

Link Analysis (Stammers, Carey & Astley, 1990; Kirwan & Ainsworth, 1992 Drury, 1995)

Link analysis represents the sequence in which device elements are used in a given task or scenario. The sequence provides the links between elements of the device interface. This may be used to determine if the current relationship between device elements is optimal in terms of the task sequence. Time data recorded on duration of attentional gaze may also be recorded in order to determine in display elements are laid out in the most efficient manner. The link data may be used to evaluate a range of alternatives before the most appropriate arrangement is accepted.

Layout Analysis (Easterby, 1984)

Layout analysis builds on link analysis to consider functional groupings of device elements. Within functional groupings, elements are sorted according to optimum trade-off of three criteria: frequency of use, sequence of use and importance of element. Both techniques (link and layout analysis) lead to suggested improvements for interface layout.

Hierarchical Task Analysis (Annett, Duncan, Stammers & Grey, 1971; Stammers & Shepherd, 1995)

Hierarchical Task Analysis (HTA) has been a technique central to the discipline of Ergonomics for over 20 years. Application of the technique breaks tasks down into goals, plans and operations in a hierarchical structure. Whilst the technique offers little more than a task description, it serves as the input into other predictive methods, for example PHEA and KLM. The concepts of HTA are relatively straightforward, but the approach requires some practice and reiteration before HTA can be applied with confidence.

Predictive Human Error Analysis (Embrey, 1993; Stanton, 1995; Baber & Stanton, 1996b)

Predictive Human Error Analysis (PHEA) is a semi-structured human error identification technique. It is based upon Hierarchical Task Analysis (HTA) and an

error taxonomy. Briefly, each task step in HTA is taken in turn and potential error modes associated with that activity are identified. From this the consequences of those errors are determined. PHEA appears to offer reasonable predictions of performance but may have some limitations in its comprehensiveness and generalisability.

Repertory Grids (Kelly, 1955; Baber, 1996)

Repertory grids may be used to determine people's perception of a device. In essence, the procedure requires the analyst to determine the elements (the forms of the product) and the constructs (the aspects of the product that are important to its operation). Each version of the product is then rated against each construct. This approach seems to offer a way of gaining insight into consumer perception of the device, but does not necessarily offer predictive information.

Keystroke Level Model (Card, Moran & Newell, 1983)

The Keystroke Level Model (KLM) is a technique that is used to predict task performance time for error-free operation of a device. The technique works by breaking tasks down into component activities, e.g. mental operations, motor operations and device operations, then determining response times for each of these operations and summing them. The resultant value is the estimated performance time for the whole operation. Whilst there are some obvious limitations to this approach (such as the analysis of cognitive operations) and some ambiguity in determining the number of mental operations to be included in the equation, the approach does appear to have some support.

Summary of Approaches

As can be seen there is an immense variety in the range of approaches in what they address (i.e. the human element, the device element or the interaction) and what they produce (e.g. task descriptions, predicted errors, performance times). For the purposes of considering where in the design life-cycle each of the methods was most appropriate, we summarised design to six main phases, namely:

Concept: in which the idea for the device is considered in a largely informal manner, many implementations are considered and many degrees of freedom remain.

Flowsheeting: in which the ideas for the device become formalised and the alternatives considered become very limited.

Design: in which the design solution becomes crystallised and blueprints are devised.

Prototyping: in which a hard built prototype device is developed for evaluation.

Commissioning: in which the final design solution is implemented and enters the marketplace.

Operation and maintenance: where the device is supported in the marketplace.

As table five suggests, there are differences in the methods in terms of the design stage they can be used, the learning and application time, their ease of use, the relationship between advantages and disadvantages, their usefulness and overall assessment. In our opinion, link analysis, layout analysis and PHEA offered the best overall utility.

TABLE FIVE ABOUT HERE

Table 5. Authors' summary of methods reviewed (on a rating scale of 1 to 5 where 1 is poor and 5 is good)

These methods were be evaluated in a training context, which is reported in the next section. It would be interesting to examine the difference in assessment by novice users and expert users.

Training people to use ergonomics methods

Very little has been written on training people to use ergonomics methods as noted by Stanton & Stevenage (1996). In order to evaluate the ease with which people are able to acquire ergonomics methods, we decided to conduct a study into the training and application by novice analysts. For the purpose of this study we used the 11 methods outlined in section 3.

Participants

Eight male participants and one female participant were recruited from the Faculty of Engineering at the University of Southampton. The age range of participants was from 19 to 25 years.

Design

A completely repeated factorial design was utilised so that participants experienced all methods in the training, practice and application sessions.

Procedure

The procedure contained two main phases, training in methods (in the first week) and application of the methods (in the second week) to the evaluation of a device. These were as follows.

Training session in ergonomics methods

In the first week, participants spent up to a maximum of four hours per method, including time for practice. The training was based upon tutorial notes developed for training ergonomics methods by one of the authors. The training for each method consisted of an introduction to the main principles, an example of applying the method by case study, and the opportunity to practice applying the method on a simple device. In order to be consistent with other training regimes in ergonomics methods, the participants were split into small groups. In this way they were able to use each other for the interviews, observations, etc. At the end of the practice session each group presented their results back to the whole group and experiences were shared. Timings were recorded for training and practice sessions.

Test session applying ergonomics methods

In the second week participants applied each method in turn to the device under analysis. Timings were taken for each method and subjective responses to the methods were recorded on a questionnaire. Analysis of the validity of the participants' responses are not included as this is beyond the scope of this paper, here we assess the usability of the methods.

Following the test session, participants were thanked for their time and paid for participating in the study.

Materials

An ergonomics methods training manual was developed to train participants and was accompanied by overhead transparencies during the training session. Participants were allowed to use the manual during the application sessions. For the purpose of applying the methods to the evaluation of a device, nine radio-cassette machines (Ford 7000 RDS EON). Timings were taken from stopwatches and participant evaluation of each

method was recorded by questionnaire. The questionnaire contained seven questions asking them to make judgements about the following criteria from Kirwan (1992):

Acceptability: the overall degree to which participants find the process and outcome acceptable, this is analogous to content and face validity.

Auditability: the degree to which participants feel that the method and its output are open to external scrutiny.

Comprehensiveness: the breadth of coverage of the technique and the extent to which it is able to describe a range of behaviour, this is analogous to concurrent and predictive validity.

Consistency: the degree to which the method is likely to produce the same result on successive occasions, this is analogous to test-retest reliability.

Theoretical validity: the extent to which the method has theoretical foundations.

Resource usage: the amount of resources, usually time and effort, required to conduct the evaluation with a given method.

Usefulness: the degree to which the participants found the method to offer a useful output.

These criteria provide a basis for comparing the subjective evaluation of the methods along a seven point Likert scale, indicating that a rating of '7' is always 'good' and a rating of '1' is always 'poor'.

Analysis

All data were analysed using the Friedman Two-Way ANOVA. Visual inspection of the homogeneity of variance for the time data confirmed that non-parametric analyses were appropriate. A correction for the multiple tests was applied to ensure no type 1 errors.

Results and Discussion of Training Study

The data from the training and practice phase do not lend themselves to statistical analysis because they were taken for the group as a whole. This can be justified by virtue of the fact that most formal training sessions in ergonomics methods occurs within a group context. Nevertheless, the data do present an interesting picture, as

shown in figure 2. These data seem to reinforce the reason for the popularity of questionnaires, interviews, observations, checklists and heuristics noted in the survey as they take relatively little time to learn when compared with HTA and PHEA. Perhaps it is surprising to see that link and layout analysis are not more popular given that they are also relatively quick to train people in. Similarly repertory grids and the keystroke level model seem to be no more time consuming to train people in than the focused interview. However, these techniques are rather more specialised in their output, like link and layout analysis.

FIGURE TWO ABOUT HERE

Figure 2. Training and practice times for ergonomics methods.

The picture for application of the methods is rather similar, as figure 3 shows. There is a statistically significant difference in the time taken to analyse a device using different approaches (Chi-square = 69.1061; $p < 0.0001$). We did not compute comparisons between individual methods because the non-parametric tests were not powerful enough to cope with the small sample size and the large number of ties in the data. Thus, as the overall ANOVA was statistically significant, it was deemed that a visual inspection of the results was sufficient.

FIGURE THREE ABOUT HERE

Figure 3. Average times for execution of each method.

As figure 3 shows, the popularity of questionnaires, observations and checklists is reinforced by these being relatively quick and flexible methods. It is worth noting that heuristics and interviews appear to take as long as link analysis, repertory grids and the keystroke level model, whereas layout analysis appears quicker. HTA and PHEA, whilst taking approximately the same as each other are substantially more time intensive than other methods. It is also worth noting that PHEA requires the output of HTA and therefore this technique would require the time to conduct HTA plus the time to conduct PHEA if it was to be used in a situation where no HTA had been developed.

The subjective evaluation of the methods by the participants over the seven criteria also produced some interesting results. Acceptability is presented because, although these data were not statistically significant, it does provide an overall impression of the participants' preferences for methods. Overall acceptability of the methods again

indicates a strong preference for interviews, observation and heuristics, as shown in figure 4.

FIGURE FOUR ABOUT HERE

Figure 4. Average ratings for acceptability for each method (on a rating scale of 1 to 7 where 1 is poor and 7 is good)

Figure 4 also shows that layout analysis, HTA, repertory grids, KLM and questionnaires are quite highly rated by participants. The surprising result is the poor rating for checklists (achieving an overall rating similar to link analysis), but this could be due to the particular applicability of the checklist used in this study. Only two statistically significant findings were found in the subjective evaluations, these were for the consistency of the methods and the resource usage. Participants rate some methods as significantly less consistent than others (Chi-square = 39.6061; $p < 0.0001$), as shown in figure 5.

FIGURE FIVE ABOUT HERE

Figure 5. Average ratings for consistency (on a rating scale of 1 to 7 where 1 is poor and 7 is good)

As figure 5 shows, heuristics are rated as less consistent than any other method, whereas more structured techniques (e.g. checklists, HTA, PHEA and KLM) are rated as more consistent. It is ironic, but not surprising, that the highest rated method in terms of consistency was also rated as one of the least acceptable methods. Some methods were also rated as requiring significantly more resources than other methods (Chi-square = 37.6869; $p < 0.0001$) as shown in figure 6. This analysis seems to favour questionnaires, checklists, observation, repertory grids and KLM. HTA is obviously resource intensive, as are PHEA, link analysis and interviews.

FIGURE SIX ABOUT HERE

Figure 6. Average ratings of resource usage (on a rating scale of 1 to 7 where 1 is poor and 7 is good)

These analyses seem to suggest that some methods will be more acceptable than others because of the time required to learn to use them, the time they take to apply to an

evaluation and the degree of consistency that they offer. Implications of this study together with the survey will be considered in the general discussion.

General Discussion and Conclusions

The general aim of this paper was to attempt to answer four questions posed in the introductory section, these question serve to structure the general discussion. It should be noted that the methods referred to in this study are all about previously developed instruments.

Which method is appropriate?

The appropriateness of the application of ergonomics methods to points in the design life cycle of products and devices is one of continuing debate. In table 5 we attempt to identify stages that we perceive the benefit to be optimal. Obviously some methods depend upon the existence of a device to evaluate (such as observation, link analysis and layout analysis) whereas others do not (such as heuristics, checklists and repertory grids). An interesting picture is painted by the survey which asks what people use the methods for. The responses showed that four main areas of application were highlighted: data collection, design, assessment and validation activities. Table 6 summarises the methods used in these general areas. The findings agree with our assessment of interviews and repertory grids, i.e. that they are appropriate for most of the design stages.

TABLE SIX ABOUT HERE

Table 6. Application of ergonomics methods.

In addition, some interesting links between methods came from the survey which are shown in figure 7.

FIGURE SEVEN ABOUT HERE

Figure 7. Links between methods.

As shown, the interview is directly and indirectly linked to five other methods. This makes the interview an important design method. Given the concern about reliability and validity of the interview in other fields of research (Cook, 1988) we would caution users of this technique to ensure that they employ a semi-structured and situationally focused approach to the device evaluation interview.

How long does it take to train people to use the method?

From our studies we have shown that initial training and practice time in ergonomics methods is quite varied depending upon the technique being addressed. Questionnaires are undoubtedly the quickest to train and practice whilst HTA and PHEA undoubtedly take the longest time of the methods we evaluated. Table 7 provides a rough guide to training times for comparison purposes. We have to point out that this is the first study conducted and therefore exact time values must be treated with caution. They do, however, provide the reader with an approximation of the relative differences between the methods.

TABLE SEVEN ABOUT HERE

Table 7. Comparison of combined training and practice times for ergonomics methods.

As this study trained participants in small groups, individual analyses of performance to some predetermined criterion was not possible. This has obvious methodological limitations for the research, but we accept these limitations within the applied nature of the research project. We would like to suggest that these issues should be addressed in future research, however.

How long do people take in applying the method to evaluate a device?

In a similar vein to the last question, application times varied between methods. Again, the questionnaire was undoubtedly the quickest to apply whilst HTA and PHEA undoubtedly took longer to apply in the device evaluation study. The only difference in the time categorisation in comparing tables 7 and 8 were for the two methods which took longer to apply than to train and practice, i.e. heuristics and link analysis.

TABLE EIGHT ABOUT HERE

Table 8. Comparison of applications times for ergonomics methods.

What are the relative benefits of one method over other methods?

In assessing the relative benefits of one method we can consider the applicability of the approaches (which would favour interviews and repertory grid as generic approaches) and the training (which would favour the application of the questionnaire as a quick approach) and application (which would again favour the application of the questionnaire as a quick approach) times. In addition we assessed the subjective

evaluation of the people who used the methods in our study and the 'ease-of-use' data from our survey of professional ergonomists. The survey suggests that professional ergonomists prefer to use checklists, guidelines and interviews. Checklists were rated as the most consistent method by the people in our training study and questionnaires were rated as the least resource intensive together with KLM.

Conclusions

In conclusion, there is clearly little reported evidence in the literature of reliability or validity of ergonomics methods. This was confirmed by the survey. The patterns of usage suggest that there is no clear match of methods to applications, which presents a rather confusing picture when embarking upon an evaluation. Apart from a few clearly defined applications the pattern looks almost random. The detailed review of ergonomics methods led to a greater insight into the demands and outputs of the methods under scrutiny. The training study indicated that link analysis, layout analysis, repertory grids and KLM appear to offer good utility when compared with other, more commonly used, methods. We would suggest ergonomists and designers would be well served by exploring the utility of other methods rather than always relying upon 3 or 4 of ones favourite approaches. However, it is an important goal of future research to establish the reliability and validity of ergonomics methods. These data could provide the encouragement for designers to try alternative approaches.

ACKNOWLEDGEMENT

This research reported in this paper was supported by the LINK Transport Infrastructure and Operations Programme

References

- Annett, J., Duncan, K. D., Stammers, R. and Grey, M. J.** 1971 Task analysis. Department of Employment Training information paper 6. HMSO, London.
- Baber, C.** 1996 'Repertory grid theory and its application to product evaluation' in **P. W. Jordan; B. Thomas; B. A. Weerdmeester & I. L. McClelland** (eds) *Usability Evaluation in Industry* Taylor and Francis, London, pp 157-165
- Baber, C. and Mirza, M. G.** 1996 'Ergonomics and the evaluation of consumer products: surveys of evaluation practices' in **N. A. Stanton** (ed) *Human Factors in Consumer Product Design* Taylor and Francis, London, in preparation.
- Baber, C. and Stanton, N. A.** 1996a 'Observation as a usability method' in **P. W. Jordan; B. Thomas; B. A. Weerdmeester & I. L. McClelland** (eds) *Usability Evaluation in Industry* Taylor and Francis, London, pp 85-94

- Baber, C. and Stanton, N. A.** 1996b 'Human error identification techniques applied to public technology: predictions compared with observed use' *Applied Ergonomics* **27** (2) pp 119-131
- Brooke, J.** 1996 'SUS: a 'quick and dirty' usability scale' in **P. W. Jordan; B. Thomas; B. A. Weerdmeester & I. L. McClelland** (eds) *Usability Evaluation in Industry* Taylor and Francis, London, pp 189-194
- Card, S. K., Moran, T. P. and Newell, A.** 1983 *The Psychology of Human-Computer Interaction*. Erlbaum, Hillsdale NJ.
- Cook, M.** 1988 *Personnel Selection and Productivity*. Wiley, Chichester.
- Corlett, E. N. & Clarke, T. S.** 1995 *The Ergonomics of Workspaces and Machines*. 2nd Edition. Taylor and Francis, London.
- Diaper, D.** 1989 *Task Analysis in Human Computer Interaction*. Ellis Horwood, Chichester.
- Drury, C. G.** 1995 'Methods for direct observation of performance' in **J. Wilson and N. Corlett** (eds) *Evaluation of Human Work*. 2nd Edition. Taylor & Francis, London, pp 45-68.
- Easterby, R.** 1984 'Tasks, processes and display design' in **R. Easterby and H. Zwaga** (eds) *Information Design* Taylor and Francis, London.
- Embrey, D.** 1983 'Quantitative and qualitative prediction of human error in safety assessments' in the Institution of Chemical Engineers symposium Series 130 pp 329-350.
- Jordan, P. W.; Thomas, B.; Weerdmeester, B. A. & McClelland, I. L.** (1996) *Usability Evaluation in Industry*. Taylor and Francis, London.
- Kelly, G. A.** 1955 *The Psychology of Personal Constructs*. Norton, New York.
- Kirwan, B. and Ainsworth, L.** 1992 *A Guide to Task Analysis*. Taylor & Francis, London.
- Kirwan, B.** 1994 *A Guide to Practical Human Reliability Assessment*. Taylor and Francis, London.
- Nielsen, J.** 1992 'Finding usability problems through heuristic evaluation' in *Proceedings of the ACM Conference on Human Factors in Computing Systems* ACM Press, Monterey CA, pp 373-380.
- Ravden, S. J and Johnson, G. I.** 1989 *Evaluating Usability of Human-Computer interfaces: a practical method*. Ellis Horwood, Chichester.

- Sinclair, M.** 1995 'Subjective assessment' in **J. Wilson and N. Corlett** (eds) *Evaluation of Human Work*. 2nd Edition. Taylor & Francis, London, pp 69-100.
- Stammers, R. B. Carey, M. and Astley, J. A.** 1990 'Task analysis' in **J. Wilson and N. Corlett** (eds) *Evaluation of Human Work..* Taylor & Francis, London, pp 134-160.
- Stammers, R. B. and Shepherd, A.** (1995) 'Task analysis' in **J. Wilson and N. Corlett** (eds) *Evaluation of Human Work*. 2nd Edition. Taylor & Francis, London, pp 144-168.
- Stanton, N. A.** 1995 'Analysing worker activity: a new approach to risk assessment' *Health and Safety Bulliten*, 240 pp 9-11.
- Stanton, N. A. and Baber, C.** 1996 'Factors affecting the selection of methods and techniques prior to conducting a usability evaluation' in **P. W. Jordan; B. Thomas; B. A. Weerdmeester & I. L. McClelland** (eds) *Usability Evaluation in Industry*. Taylor and Francis, London pp 39-48.
- Stanton, N. A. and Stevenage, S.** 1996 Learning to predict human error: issues of reliability, validity and acceptability. Submitted to *Ergonomics*
- Stanton, N. A. and Young, M.** (1995) *Development of a methodology for improving safety in the operation of in-car devices* EPSRC/DOT LINK Report 1. University of Southampton, Southampton.
- Wilson, J.** 1995 A framework and context for ergonomics methodology in **J. Wilson and N. Corlett** (eds) *Evaluation of Human Work*. 2nd Edition. Taylor and Francis, London pp 1-39
- Wilson, J. and Corlett, N.** 1995 *Evaluation of Human Work*. 2nd Edition. Taylor and Francis, London.
- Woodson, W. E., Tillman, B. and Tillman, P.** 1992 *Human Factors Design Handbook* 2nd edition, McGraw-Hill, New York.
- Youngman, M. B.** 1982 *Designing and Analysing Questionnaires*. TRC, Maidenhead.

Author	Title	Edit	Date	Pages	Mthd1	Dmn	Mthd2	Vldtn	Trng
Diaper	Task Analysis in HCI	E	1989 (1st)	258	6	S	S	0	0
Kirwan & Ainsworth	A Guide to Task Analysis	E	1992 (1st)	417	23	G	G	0	0
Kirwan	A Guide to Practical HRA	A	1994	592	28	G	S	2	0
Corlett & Clarke	Ergonomics of Workspace and Machines	E	1995 (2nd)	128	6	G	G	0	0
Wilson & Corlett	Evaluation of Human Work	E	1995 (2nd)	1134	48	G	G	1	0
Jordan et al	Usability Evaluation in Industry	E	1996 (1st)	252	20	G	G	2	0

Table 1. An overview of six ergonomics' methods books.

KEY TO TABLE ONE:

Term	Definition
Author	Author(s) or Editors of text
Title	Main title of text
Edit	Authored (A) or Edited (E) book
Date	Latest year of publication with edition: 1st or 2nd.
Pages	Number of main pages in text including index
Mthd1	Number of ergonomics' methods covered by the text
Dmn	Domain of application: Specific (S) or General (G)
Mthd2	Range of methods: Specific (S) or General (G)
Vldtn	Number of validation studies cited within text
Trng	Number of training studies cited within text

Application	Method(s)
Data Collection	Questionnaires, interviews and observation
Design	Simulators, computer simulation, interviews and repertory grids
Assessment	Simulators
Validation	Repertory grids

Table 6. Application of ergonomics' methods.

Training time	Method(s)
Over 3 hours	HTA and PHEA
1 to 2 hours	Interviews, Repertory grids, KLM
30 mins to 1 hour	Heuristics, Checklists, Observations, Link analysis and Layout analysis
Less than 20 mins	Questionnaires

Table 7. Comparison of combined training and practice times for ergonomics' methods.

Application time	Method(s)
Over 3 hours	HTA and PHEA
1 to 2 hours	Heuristics, Interviews, Link analysis, Repertory Grids, KLM
30 mins to 1 hour	Checklists, Observations, Layout analysis
Less than 20 mins	Questionnaires

Table 8. Comparison of applications times for ergonomics' methods.