USABILITY — CONTEXT, FRAMEWORK, DEFINITION, DESIGN AND EVALUATION

BRIAN SHACKEL

1. Introduction

For many users the informatics system is essentially the terminal or workstation which they are using, and that is the central computer as they see it. But only too often these users are seen as 'end-users' by designers — and this name may well betray an attitude which causes some of the bad design for users and failures in usability. Designers must see the user as the centre of the computer system instead of as a mere peripheral. This simple concept, easy to state but harder to achieve, is often expounded by ergonomists and human factors specialists. It has been emphasised by Nicholls (1979):

> "In spite of changes in the nature of computing, remnants of old thinking remain with us. In former days, when the CPU was at the heart of a system, designers naturally talked of 'terminals' and 'peripherals'. I suspect it was in this period that people began to use the term 'end user'. The unconscious symbolism is both a symptom and a cause; the 'end' user at the 'terminal' was often the last person to be considered in the design of the system. It is important to develop a new view of computing systems, and to look at the user in a different light ... taking this view of computing, the centre of a system is the user."

So, if we are to improve the usability of interactive computer systems, then the former orientation of designers must be completely reversed.

2. Usability context — the acceptability equation

When users and purchasers make decisions about systems, their decision depends not only upon usability but upon an assessment balancing various factors; they probably consider also how useful the system will be, whether they feel it is suitable and they would like to use it, and how much it will cost, both financially and in terms of the personal, social and organisational consequences. Without being able as yet to define a mathematically precise relationship between these terms, it is suggested that the relevant factors are associated in some form of trade-off paradigm such as that in Figure 1. This paradigm suggests that whether I accept something depends upon whether I consider it sufficiently useful, usable and likeable in relation to what it costs me. If I do not accept something, then the combination of utility, usability and likeability are not sufficient for it to satisfy my wants in relation to human and financial costs.

UTILITY	— will it do what is needed functionally?
+	
USABILITY	— will the users actually work it successfully?
+	
LIKEABILITY	— will the users *feel* it is suitable?
must be balanced in a trade-off against	
COST	— what are the capital and running costs?
	— what are the social and organisational consequences?
to arrive at a decision about	
ACCEPTABILITY	— on balance the best possible alternative for purchase.

Figure 1: The paradigm of usability and related concepts

Thus, this paradigm helps to place usability in its balanced position with functionality; as computers become cheaper and more powerful, it seems certain that usability factors will become more and more dominant in the acceptability decisions made by users and purchasers.

3. Usability framework and criteria

Successful system design for usability requires much attention to various aspects of the user. However, the user must not be considered in isolation from other aspects of the situation; that would only be perpetuating in reverse the all too common fault in the past of considering the technological tool in isolation from the user. Good system design depends upon solving the dynamic interacting needs of the four principal components of any user-system situation: user, task, tool and environment.

Likewise usability, an important goal for good system design, depends upon the dynamic interplay of these four components (this framework is based upon earlier similar approaches by Bennett, 1972, 1979, and Eason, 1981) — see Figures 2 and 3.

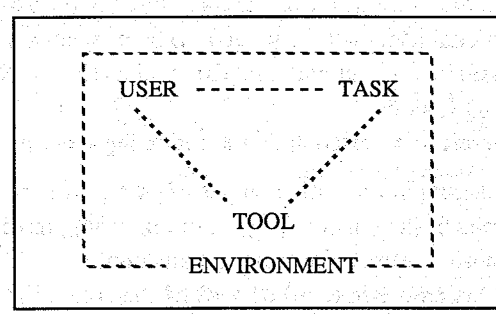

Figure 2: The four principal components in a human-machine system.

Figure 3: These joint authors as *Users* are revising a paper (*Task*) for an electronic journal using a computer and VDT (*Tool*) in the *Environment* of a research centre.

With the framework of the four principal components in mind, we can now turn to the meaning of usability. Usability depends (a) upon the design of the tool (the VDT and the computer system) in relation to the users, the tasks and the environments, and (b) upon the success of the user support provided (training, manuals, and other job aids such as on-line and off-line 'help' facilities). We consider that usability for individual users will be judged (a) by subjective assessment of ease of use of the design with its user support, and (b) by objective performance measures of effectiveness in using the tool.

Evaluation will therefore be based upon the following criteria:

- success rate in meeting the specified ranges of users, tasks and environments
- ease of use in terms of judgements (e.g., learning, using, remembering, convenience, comfort, effort, tiredness, satisfaction)
- effectiveness of human use in terms of performance (e.g., time, errors, number and sequence of activities, etc.) in learning, relearning and carrying out a representative range of operations.

4. Usability definition

From the above suggestions, it is evident that usability considered in this way is not only conceived of as *ease of use* but also equally involves *efficacy* i.e., effectiveness in terms of measures of (human) performance. Therefore, the formal definition proposed for the usability of a system or equipment is:

> 'the capability in human functional terms to be used easily and effectively by the specified range of users, given specified training and user support, to fulfil the specified range of tasks, within the specified range of environmental scenarios'.

A convenient shortened form for the definition of usability might be 'the capability to be used by humans easily and effectively', where:

easily	=	to a specified level of subjective assessment
effectively	=	to a specified level of (human) performance.

The definition of usability was probably first attempted by Miller (1971) in terms of measures for 'ease of use', and these were developed further by Bennett (1979) to describe usability. The concept of usability was first fully discussed and a detailed formal definition, as above, was attempted by Shackel (1981), and Bennett (1984) modified and developed the definition.

The problem with these definitions is that they are conceptually satisfactory but still only generalised in form; they do not specify what is usability in quantifiable or measurable terms. Therefore, I have integrated and developed further these

Proposed Operational Definition of Usability

Usability can be specified and measured by means of the operational criteria defined below. The terms should be given numerical values when the usability goals are set during the design stage of 'requirements specification'.

For a system to be usable the following must be achieved:

Effectiveness

- The required range of tasks must be accomplished at better than some required level of performance (e.g., in terms of speed and errors)
- by some required percentage of the specified target range of users
- within some required proportion of the range of usage environments

Learnability

- within some specified time from commissioning and start of user training
- based upon some specified amount of training and user support
- and within some specified relearning time each time for intermittent users

Flexibility

- with flexibility allowing adaptation to some specified percentage variation in tasks and/or environments beyond those first specified

Attitude

- and within acceptable levels of human cost in terms of tiredness, discomfort, frustration and personal effort
- so that satisfaction causes continued and enhanced usage of the system.

Figure 4: Definition of Usability proposed in terms of goals and operationalised criteria which can have numerical values specified and measured.

approaches, and now propose an operationalised definition of usability in Figure 4.

This definition has been formulated so that numerical values can be specified during the design stage of user requirements specification. In that stage of the design process, various system requirements are specified and the usability requirements should be specified in just as much detail as any other aspect of the intended system.

4.1. *Setting usability goals*

To illustrate this definition, and to demonstrate how it should be used during the design stage of requirements specification to set usability goals, the following example has been prepared.

Let us suppose that a design team is writing the specification for a software package by which either microcomputers or 'dumb' terminals on a large system may be used to dial up and login to a remote host computer, so that the user can join in a computer conference or deal with electronic mail. The required range of user tasks might be specified by the team, with the help of an ergonomist adviser, in the form shown in Figure 5. Then, using the definition of usability given in Figure 4, the desired usability goals may be set in terms of testable numerical values as shown in Figure 6.

User Requirements Specification for Electronic Mail Software Package

The required range of user tasks shall be:

1 Set parameters for 2 terminals (ADM3/VT100)
2 Set up auto-dial and auto-login

3 Initiate dial up and login
4 Start on-line usage
5 Send and receive a file
6 Set auto-record of the communication to file on/off
7 Set slave printer on/off
8 End on-line usage

9 Set up preset unattended dial up/login/file transfer
10 Read auto-recorded file
11 Read received file
12 Leave the system

Tasks 3 - 8 shall each be done by a maximum of two keystrokes.

Figure 5: Illustrative specification of user tasks as basis for Figure 6.

Thus, perhaps the most important feature of this process is that the usability goals thus set become criteria by which to test the design as it evolves and to improve it by iterative redesign. Such tests are embodied first in trials of early versions of the design and later in formal evaluations of prototypes. The bases of evaluation are discussed and some of the relevant procedures are outlined in the final sections of this paper below.

4.2 *Specifying usability attributes*

Having set usability goals at the more global level, it will usually be necessary to specify in more detail the usability attributes desired for various features of the system or aspects of its operation. A useful procedure for this process of attribute

Illustration of Defining Usability Goals

To achieve the target usability, tasks 3 - 8 shall be done:

Effectiveness

- at better than 2 seconds for each task with no more than one error per 50 attempts
- by 90% of the target range of managers, secretaries and professionals
- in any office situation which complies with the offices, shops and factories acts and which does not seriously contravene the workstation and environment specifications recommended in the V.D.T. Manual (Cakir, Hart and Stewart, 1980) or in Designing Systems for People (Damodaran, Simpson and Wilson, 1980)

Learnability

- either within 2 hours of starting use via learning with the manual and help system or within 1 hour of starting use via the training course
- based upon the 'getting started' manual to be provided or upon the 2 hour training course to be prepared
- within half-hour relearning time for users who have (1) completed the learning defined above, (2) done all the tasks once per week in half-hour sessions for 3 further weeks, and (3) then not done the tasks for an interval of 4 weeks

Flexibility

- with the task flexibility requirement only to apply to tasks 1, 2 and 9, but with the environment flexibility requirement to apply to tasks 3 - 8 that user performance shall not deteriorate more than 10% in warm conditions up to 95° F and 95% R.H.

Attitude

- and with attitude questionnaire results on 5-point scales ('very good' to 'very bad') to be at least 80% in 'good' or better and only 2% below 'neutral'
- so that questionnaire results also give at least 90% 'yes' answers to the question: "Imagine you must use this system 5 times per day every day — would you be satisfied to continue using it, and if not please comment why not?"

Figure 6: Setting usability goals via the definition of usability

specification was developed in 1978 by Gilb in his "design by objectives" methodology (later published in Gilb, 1988).

This procedure has been adapted by Bennett (1984), whence the following example and explanation are taken — see Figure 7.

"Attributes are given brief descriptions and then further defined by a series of parameter categories across the top of the table. Each attribute takes on real meaning only when we specify how we will measure it in the 'Unit of Measure' column. In the example in Figure 7, the Unit-of-Measure (hours) is chosen to be long enough to give an idea of steady-state user performance, yet short enough for feasible testing. The rest of the Parameter Categories in the sample table are used to specify values that establish various levels of performance. The figure for the Planned Value indicates a satisfactory performance for the final user of the system, and it should be consistent with the Planned Values for all other system attributes. The Worst Case Value marks the borderline between a tolerable and an intolerable system. If the observed value for any one attribute goes beyond the assigned Worst Case level for that attribute, then the system as a whole may be formally unacceptable. While the system may not be 'worthless', it has not met the goals. An explicit Current Value provides a basis for comparisons. For example, management may be prepared to accept a lower-than-currently-available level for user performance for part of a system if the overall level of system performance, as measured by other high-priority attributes, is greater for the new system. The Remarks column may contain a reference to source information or to additional details listed in another table containing an expansion for this attribute.

In Figure 7, Attribute *a* relates to learnability. Project team members specify that the 'easy to learn' attribute will be measured by five people using a prototype mockup to perform a sample teleconferencing task. For a particular scenario, the estimated Planned Value is four hours. A refined measure would give an estimate of expected deviation. The Worst Case Level provides an upper bound value; all agree that something is seriously wrong if, on the average, it takes the user more than six hours to learn the task and perform to the standard. The basis for this measure is not given here, but of course it would have to be available. The table shows that there is no Current Value data available to show the time for learning comparable performance skills with an alternative conferencing tool.

Attributes *b*, *c*, and *d* focus on throughput. Attribute *e* addresses one measure of flexibility as the new teleconferencing function is integrated into people's work patterns. Attribute *f* shows how attitude will be measured. Note that this table only gives an overview at a glance. The details for Attributes and for Parameter Categories must be contained in supporting documents."

This method raises three questions.

First, how do we set the values shown in each parameter category? At this stage in the development of these concepts, most requirements people are not accustomed to supplying goals specific enough to make careful trade-offs. In addition, development people are not accustomed to receiving such goals. For these reasons,

Development Goals for Usability of a System

PARAMETER CATEGORIES

Attributes	Means Used to Measure	Unit of Measure	Planned Value	Worst Case Value	Current Value	Remarks
a. Basic Conf. tool must be easy to learn	test sample of 5 people using prototype	hours, score	4	6	no data	must be able to learn
b. Conf. tool leads to results comparable to face-face	sample task suitable for tele-conf.	hours to complete structured interview	2	4	4	a key selling point
c. Relation of errors in Conf. use to errors in other parts	kind and rate of errors	count of errors due to confusion about design	1/hr	5/hr	?	must fit existing style of system use
d. Recovery from errors (system or user)	observe log of user reactions	count of errors requiring > 30 sec.	hold to 2/hr	more than 7/hr	1/hr in current system	field reps say is of key import
e. Smooth transitions from menus to commands in use of new function	task set requiring use of new and old function	count actions, time, to make transitions	less than 1/10 overall time	1/3	1/10 time is currently spent overall moving between current system parts	
f. Attitude toward continued use	interview after questionnaire	score	80%	50%	current system score 80%	word-of-mouth is key for referral selling

Format adapted from Gilb (1988)

Figure 7: A sample attribute specification table showing parameters for usability attributes. Each objective is established by the values chosen for that parameter category. (Quoted with permission from Bennett, 1984.)

values must be arrived at through iteration and negotiation. At the least, this method places important values "on the record", and thus avoids unpleasant surprises — such as requirements people envisioning an on-the-job training time of three hours and development people assuming that users will attend a 30-hour class.

Second, how do developers make trade-offs intended to meet goals? Again, there is no magic process, but both the target values and the trade-offs made are explicit in this approach. In a private communication, Stuart Card has pointed out that making trade-off points explicit can have a beneficial effect on use of system resources. For example, it may be very expensive to achieve three-second response time for a particular kind of information retrieval request. If an analysis of goals with respect to work patterns reveals that such a request comes at a natural closure point for users, then the response time can be reasonably adjusted upward for that requirement without damaging the overall usability of the system. If target values and trade-off points are explicit, then there is hope for tracing results observed in the field back to specific design decisions so that we can learn from experience.

Third, how often must progress be monitored? The scope and frequency of monitoring the process is a business decision. For example, a minor new release of standard technology for an existing product intended for training specialists is usually less tricky than introducing a product intended for a new user group.

Among other issues which may be raised is the suggestion that parts of this process are impossible; for example, some attributes may be considered "unmeasurable". But, as Gilb states, if the existence of a quality can be determined, then it is measurable — if only at the level of "present" or "absent". It is better to have some measure for an important system quality (even if the method to measure it is weak) than to have no measure whatsoever. No measure means no hope of control over that particular quality. If there is no convenient objective measure, then we can make use of sampling and statistical methods from the social sciences in order to quantify opinion.

Finally, there is the question of who should be directly involved with setting goals and specifying attributes. Too easily this may be seen solely as the work of the designers (not users) and of the technologists (not managers). On the contrary, to achieve usability in the ultimate design it is essential that users and managers are fully involved in this specification process (as is discussed in the next section of this chapter). Gilb reports that one top manager's initial reaction to an Attribute Specification Table was, "That's for the technical people." Gilb countered with, "No, the set of tables is your primary instrument of control." It is not good sense to launch a costly project without getting all members of the team to reach a clear agreement about what they are trying to accomplish.

5. Usability design – process and precepts

The place of human factors in relation to various stages of the design process, and the best procedures for assisting designers to achieve good usability design, have been studied intuitively and empirically for many years. Meister and Farr (1967) showed some of the problems designers have in utilising human factors information; various handbooks and textbooks have been produced with a focus on general ergonomics in relation to general systems (McCormick, 1976, 1982; Van Cott and Kincade, 1972), and more recently some handbooks on the human factors of computer systems have appeared (notably Shneiderman, 1987, and Helander, 1988). However, relatively few attempts have been made to give prescriptive advice on how to bring ergonomics into the design process (cf. Christensen, 1971; Shackel, 1971, 1974 Chapter 2); moreover, it is only recently that an attempt has been made to do this in relation to the design of computer systems (cf. Damodaran *et al.*, 1980 and the chapters by Damodaran, Eason and Harker, and Gardner in this book).

However, in the last few years two strands of development appear to be converging towards a common set of precepts. Based upon a wide range of research and design experience at the HUSAT Research Centre, Eason (1982, 1983) described various issues involved in the process of introducing information technology, and proposed an evolutionary system development process; this includes various ways of involving examples of the users, pilot systems, trials and experiments, progressive implementation of facilities, evaluation of users, user support and assistance to help the learning by the organisation. Some of these procedures are reported by Miller and Pew (1981) as being used by them in the course of a large system development study. Moreover, Gould and Lewis (1983) have similarly devised a methodology from their experience and have proposed four precepts for design for usability which in essence are very similar; they also give examples of the use of simulation and prototyping as part of the usability development process.

From these various approaches, one can synthesise and propose a set of fundamental features which will probably find widespread acceptance by experienced human factors specialists as key precepts for the process of design for usability. These are listed in Figure 8 below. The essentials of these fundamental features are as follows:

User-Centred Design – designers must understand who the users will be and what tasks they will do. This requires direct contact with users at their place of work. If possible, designers should learn to do some or all of the users' tasks. Such studies of the users must take place before the system design work starts, and design for usability must start by creating a usability specification. (See also Norman and Draper, 1986).

Participative Design – a panel of expected users (e.g., secretaries, managers) should work closely with the design team, especially during the early formulation stages and especially when creating the usability specification. To enable these users to make useful contributions, they will need to be shown a range of possibilities and alternatives by means of mock-ups and simulations. A valuable procedure, although not easy, is to write the parts of the operating manual describing the interface and how to use it; then user tests of a drawing of the interface with this draft manual can reveal potential problems before they have been embedded into the design.

Experimental Design – early in the development process the expected users should do pilot trials and then subsequently use the simulations, and later the prototypes, to do real work. Whenever possible alternative versions of important features and interfaces should be simulated or prototyped for evaluation by comparative testing. These studies should be formal and empirical, with measures of the performance and the subjective reactions of the users. Thus, ease of learning and ease of use can be assessed and difficulties revealed.

Iterative Design – the difficulties revealed in user tests must be remedied by re-design, so the cycle — design, test and measure, re-design — must be repeated as often as is necessary until the usability specification is satisfied.

User Supportive Design – this area is often left until a very late stage in the design process, and then some documentation and 'help' screens are written in a hurry at the last minute; the other aspects of user support are usually left to others by the designers, who are often unaware of their relevance and importance. Careful attention to all these support facilities can significantly assist usability (cf. Damodaran, 1986).

Relating these key precepts to the typical stages of the system design process can

Five Fundamental Features of Design for Usability

1. User Centred Design	—	focussed from the start on Users and Tasks
2. Participative Design	—	with Users as members of the design team
3. Experimental Design	—	with formal user tests of usability in pilot trials, simulations and full prototype evaluations
4. Iterative Design	—	design, test and measure, and redesign as a regular cycle until results satisfy the usability specification
5. User Supportive Design	—	training, selection (when appropriate) manuals, quick reference cards, aid to 'local experts', 'help' systems, e.g., on-line : context specific help off-line : 'hot-line' phone service

Figure 8: To be successful, Design for Usability must be based upon these Five Fundamental Features.

provide both a first level of elaboration of the precepts and a reminder of the action programmes required. An outline of the usability actions appropriate to the system design stages is given in Figure 9.

It should be noted that these precepts (Figure 8) are derived from separate groups, one at a university research centre and the other at the largest computer corporation.

Usability Actions in the Stages of System Design

System Design Stage	Usability Actions
Feasibility	*Define* range of *Users*, *Tasks* and *Environments* to be covered. Do the proposals match the needs? Preliminary functions and operations analyses. Preliminary allocation of functions. *Participative Design* – panel of users in the design team. Create and formalise the *Usability Specification* by defining *user requirements* and setting *usability goals*.
Research	*Studies*, often experimental, of *human capabilities* re system operational concepts. Use *pilot studies* in the field to explore *users' operational needs* and to study possible effects upon *organisational and social structure*.
Development	*Detailed analyses* of all functions, tasks and operations involving or affecting humans. *Design* all human factors aspects of equipment and workplaces. *Specify* all environmental issues. Use guidelines to assist as design ideas are developing. *Check* design ideas against available human dimension, behaviour and performance data. *Test* subsystem sections in initial *evaluation trials* with samples of likely users. *Iterative Design* – use test results as basis for redesign, and test again. Propose *selection* criteria (if relevant); develop *training* scheme; provide for other forms of *User Support* needed.
Prototype	Extensive *laboratory evaluation* with samples of likely users. *Full field trials* with representative actual users in proper working environment. *Iterative design.*
Regular Operation	Provide for *User Support* – provide training, encourage and aid 'local experts', arrange 'hot-line' for help, etc. Gather extensive evaluation data (both objective performance data and subjective attitude data); *feed back* the evaluation data as check on decisions and predictions made during design; learn from the data – *modify the design databases, models and methods* for future use.

Figure 9: A synopsis of the various activities needed in the successive stages of system design to improve usability.

The various precepts have been recommended separately by many human factors specialists but none before have integrated them in this way and shown their value as a totality from practical examples.

Further, three studies reported at the CHI '85 conference all add illustrative support to the usability specification process and design procedures outlined above. Wilson and Whiteside (1985) show the practicality and advantage of specifying usability metrics and formally defining usability goals. Olson (1985) illustrates the benefit of deliberately designing alternative versions for each part of the user interface aspects of a first prototype. Butler (1985) presents a case study to emphasise the value of setting usability goals and measures at the requirements stage, and he also describes the process as 'evolutionary design'.

6. Usability evaluation

6.1 *Evaluation bases*

Evaluation is an important topic. Chapanis (1981) has reviewed the needs and basic procedures. Hirsch (1981) described the work and procedures of the IBM San Jose human factors centre, which does many evaluation studies. Neal and Simons (1983) described a very useful recording and playback facility used at that same centre. Grudin and MacLean (1984) described various methods for measuring performance and preference, Helmreich (1984) presented the results of user acceptance research and Raveden and Johnson (1989) have proposed a substantive checklist approach to evaluation. In his chapter on 'Evaluating Usability' in this book Chapanis brings together many of the relevant issues.

In previous sections usability has been discussed in terms of its four major components (user, task, tool and environment), and has been shown to depend upon the interaction of these four major components of the system situation. This leads to the question: is it possible, reasonable and measurable to relate the concept of usability to, for example, a specified population being able to achieve some specified performance level? This, it will be remembered, was part of the definition proposed.

The first answer must be to pose a counter question — how else can the product be designed for a market with some reliable and rational (i.e., scientifically based) assurance of success? The second answer is that this has already been done, and if the problem can be solved for one type of product then surely it can be solved for others. As an example consider the US legislation (Federal Register, 1971) and the British Standard (1975) on childproof medicine containers, both of which are essentially concerned with performance and not material or dimensional issues; they specify that at least 85% of a test panel of children shall be unable to open the containers before a demonstration, 80% still unable after a demonstration, and at least 90% of a panel of adults shall be able to open and properly reclose them following written instructions only.

There are three general types of measurement available for evaluation: dimension, performance and attitude — see Figure 10. *Dimensional criteria* are the most familiar and simplest, relying on physical measurement; the same procedures are involved for human usability, but primarily in relation to human body size. The problem with analytic dimensional criteria is that they do not enable judgement that something is more useful simply because it is 2" higher, etc.; ultimately dimensions must be related to other criteria based upon human performance or attitude if any scale is to be derived. In summary, dimensional criteria only allow pass/fail judgements; satisfying them may be a necessary but not a sufficient measure of usability.

Performance criteria involve an objective statement of some achievement, often in terms of time and errors, against which human performance can then be measured. Although the interpretation of performance criteria for evaluation purposes is often also in terms of pass or fail, the measurements obtained for comparison with the criterion give some indication of the degree of usability achieved.

Attitude criteria can be defined with the same precision and operational form as performance criteria. There has been much research in psychology on controlled methods of gathering subjective data from humans, and various forms of scaling technique are now well developed and proven.

It must be emphasised that these three types of criteria and measurement should not be regarded as alternatives, but as complementary, with regard to the assessment of usability. This is perhaps evident from the fact that different types of measurement are involved, which clearly will assess different characteristics of the tool, along with the task and environment, in relation to the user.

Dimensions will be primarily relevant to the size, shape and other characteristics of the tool in relation to human size and related requirements. *Performance* will assess the operational capability which can be achieved by the human user, but of course will not assess the cost or difficulty for the user. The *attitude* measures assess the user's view of the cost and relative difficulty in achieving the performance. We should note that attitude criteria are no less valid than any other; indeed in many respects they are more valid with regard to usability, because ultimately it is the human user who must express the judgement of this characteristic. Performance measures cannot be the sole criterion, because the human may readily achieve a given performance, but still not prefer to do the task or use the tool because it is very inconvenient and awkward, so that he may well prefer (i.e., find more usable) another similar tool which gives less speed or more errors but is easier or more convenient.

Criteria	**Types of Measurement**
Dimension (analytic)	physical; anthropometric
Performance ('objective')	physiological; operational; experimental; functional
Attitude ('subjective')	psychological, e.g., by controlled scaling techniques

Figure 10. The general types of criteria and measurement available for evaluation.

6.2 *Evaluation procedures*

The above discussion attempts to provide a simple analytical framework for the issues of criteria and measurement in relation to usability. The procedures involved in system evaluations during design and after installation both include and re-orient the above into appropriate operational and time-scheduled processes. In many respects, the processes used for the human factors evaluation of system usability are similar to those used for engineering evaluation of system utility. Some brief comments only will be made on a few points of relevance.

There are three principal evaluation procedures used in human factors:

1. *Expert Review*: appraisals by human factors specialists, using the measures of dimensions (and other analytic comparison data), and of attitude (by 'expert opinion').
2. *Simulation Trials*: experiments with mock-ups and prototypes, with limited number of subjects but essentially equivalent to ultimate users, using measures of performance and attitude.
3. *User Performance Tests*: full experimental studies of final equipment with samples of actual users, using measures of dimensions, performance and attitude.

For guidance on principles and procedures, see Chapanis (1959), Meister and Rabideau (1965), Parsons (1972), Chapanis (1981) and the chapter by Chapanis in this book. While these are invaluable reference sources, we should note that on the one hand they expound basic methodology, which is very necessary, but on the other hand that their applications frame of reference mainly relates to larger military systems; there is still much to be done in modifying, developing and testing usability evaluation procedures for human-computer interaction in non-military systems. For the present, we shall recommend only one precept, which is well founded on considerable experience: attitude assessments are most reliable when users have actual 'hands-on' experience in the situation concerned, so that adequate experience (often accompanied by appropriate performance tests) is the essential pre-requisite for valid attitude measurement.

7. Conclusion

From this discussion it is evident that neither the specification of usability nor its evaluation are sufficient on their own; both must be done thoroughly and skillfully if good design for usability is to be achieved. Only in that way will the interfaces become not bottlenecks but gateways, through which the informatics system successfully interacts with and serves the user.