CHAPTER 2

METHODOLOGY FOR THE EVALUATION OF *SA GOVERNMENT ONLINE*

1. INTRODUCTION

"Evaluation is concerned with gathering information about the usability (or potential usability) of a system in order to improve features within an interface and its supporting material or to assess a complete interface" (Preece, 1993:108). A sound methodological approach is therefore necessary to gather reliable and valid data that can be used for the evaluation and subsequent improvement of a website. The evaluator should ensure that the evaluation is based on criteria that are comprehensive enough to reach findings in line with the purpose and objectives of the specific evaluation project, and that these criteria are suitable for the particular website. Similarly, it is important to select a method or methods that are suitable to assess the specific website according to the objectives of the evaluation project. The end-result of the evaluation should be to address different parts of the usability engineering life cycle (Nielsen, 1993:223) and all aspects of the website and its interface that were identified or evaluation.

To collect relevant data concerning the interface of *SA Government Online* and users' attitudes towards it, an evaluation methodology was developed by following three main steps: the identification of applicable assessment criteria or guidelines to be used, the selection of evaluation methods, and the development of test instruments for data collection. Evaluation methods were selected to be relevant to the goals of the evaluation, namely to determine if the website complied to content requirements, and to collect feedback on whether users could find the information they required, which aspects of the interface were good and bad, where the major difficulties were, and how the design could be improved.

14

2. IDENTIFICATION OF EVALUATION CRITERIA

2.1 Introduction

The first step in the development of a methodology for the evaluation of *SA Government Online* was to identify applicable assessment criteria to be used for the different test instruments selected for data collection.

According to Nielsen (1993:91,92), different levels of guidelines can be used for (heuristic) evaluation, namely general guidelines applicable to all user interfaces, category-specific guidelines for the kind of system being developed, and product-specific guidelines for the individual product. In the development of criteria for the evaluation of *SA Government Online*, all three levels were considered.

According to Ciolek (1996), evaluative activities seem to form two main streams: (a) individual work on creation of checklists or "toolboxes" of criteria that enable web information resources to be assessed, and (b) commercial, long-term projects aimed at the periodical reviews and grading of large volumes of online material. In the first case, the emphasis is on finding how the overall quality of the networked resources can be meaningfully discerned, analysed and compared. In the second case, the emphasis is on a quick separation of potentially popular materials from the rest of the web so that a site "providing such rudimentary 'filtering' services can attract Internauts and draw them towards the site's fee-based operations" (Ciolek, 1996). In developing assessment criteria, the first approach was adopted.

Identification of criteria for the evaluation of *SA Government Online* was done in the three phases that will be discussed in the following paragraphs.

2.2 Phase one: Identification of general and category-specific criteria

A wide range of sources representing the first approach was consulted, i.e. sources that attempt to specify and enumerate the essential ingredients or features of a good or quality website or sources that provide principles for user interface design. Among these were compiled lists, periodicals, journals and other sources dealing with websites, on-line rating services, and guidelines for the design and construction of high quality web pages. This process yielded a preliminary list of evaluation criteria and indicators. It should be noted

that these resources were consulted before the evaluation of the website and thus reflect dates prior to the evaluation.

2.3 Phase two: Consolidation

In the next phase the criteria and indicators were edited, synthesised and consolidated. The items were analysed to eliminate duplication and to clarify those with ambiguous meaning. Each existing list uses its own vocabulary and therefore an attempt was made to identify language for common criteria so that a diverse audience may understand it. Any item that was meaningless, which could not be operationally defined, or which did not seem relevant to the type of website being evaluated was eliminated. The relevant items were then grouped in five broad categories and 17 sub-categories to yield a standard criteria list against which all websites can be measured. Each sub-category presents a series of more detailed indicators that may assist in the evaluation process.

The criteria list includes general evaluation criteria that address substance, usability, and appearance of websites. They take into account and provide latitude in interpreting subjective elements, while maintaining a consistent standard against which all websites can be measured and which may have universal applicability for government websites. Because the sources that were studied provide criteria from the perspective of the utilisation of websites as well as the designing thereof, the criteria can be used to assist users to evaluate the quality of information on existing websites, as well as for standards to guide the design of websites. It is worth noting that this set of criteria does not necessarily cover the full spectrum of website design, but does include those which address the research problem. A detailed discussion and consolidated list of the evaluation criteria is found in chapter four.

2.4 Phase three: Identification and consolidation of category and product-specific criteria

The third phase entailed the identification of guidelines specific to the product being evaluated, i.e. the *SA Government Online* website as a government website and as a gateway/portal to other South African government websites.

Few sources provide specific criteria or guidelines for the development or evaluation of government websites, and specific guidelines or policies for the development of South African government websites did not exist at the time of this research. It was therefore

decided to provide a set of particular requirements or guidelines to be used in conjunction with the general criteria.

To determine criteria and guidelines for South African government websites some government policy documents were studied to determine the South African government's requirements and specifications for electronic information dissemination (related to web publishing) and web publishing in particular. A literature study was also conducted to identify general criteria and guidelines for government websites. Sources that discuss, describe or assess specific government websites were utilised for this study. Deductions about aspects that distinguish between good and bad sites were then made.

Guidelines selected from these sources understandably did not yield comprehensive information on usability issues, but mainly addressed content issues. They enabled and strengthened the objective and thorough evaluation of breadth and depth of information, and also contributed to determine the applicability of a website for its intended target audience.

A discussion of these guidelines can be found in chapter five.

2.5 Methodological problems experienced during criteria selection

The following methodological limitations, some of which are described in the sources consulted, were taken into account during the criteria selection:

- The weakness of guidelines is that they often generalise across a wide range of characteristics of users, tasks and environments, and it is very difficult to "rigorously specify the limits of the context in which a guideline is applicable" (Bevan et al, 1991:3). According to Ciolek (1996), concepts such as 'ease of access', 'user-friendliness', 'crisp layout' or 'detailed metadata' seem to be applied in a very general fashion, as if all documents and all resources were written in the same natural language, had the same complexity, same structure, and served the same purpose. "Can one really use the same vague, impressionistic concept to compare a single document with a collection of research papers, and finally, with a large-scale electronic archive?" (Ciolek, 1996).

- The operational meaning of criteria is ambiguous: "Does the notion of 'workability' refer to the same phenomenon identified as 'ease of finding, ease of access, good formatting and presentation' or that referred to as the 'stability of information' and 'ease of use'.

Also, how does one go about measuring the breadth, depth or thoroughness of information?" (Ciolek, 1996).

- Criteria are grouped in different ways, using own classification systems or groupings thereof.

- There are different types of websites, each with its own purpose and intended audience. When an informational website or government website is evaluated different criteria will be applied than for an entertainment website; when an information gateway is evaluated specific criteria which do not apply to a website carrying its own information content will be considered.

2.6 Operationalisation of criteria

The steps described in this section provided a complete set of evaluation criteria, which were then used in the data gathering methods selected for this research and for a critical evaluation of the website.

3. SELECTION OF EVALUATION METHODS

The process to decide on an appropriate methodology included a literature study of Nielsen (various sources), Preece (1993), Macleod (1994), Clausen (1999), Bevan et al (1991) and Gordon (2000). From these sources, the author concluded that the application of a single method to the exclusion of others would not provide complete answers to the research problem, as important aspects might be missed.

Clausen (1999:83) recommends applying more than one method. According to him at least three groups of methods are used for the design and construction of high quality websites, namely automatic procedures, exclusively quantitative methods and qualitative/heuristics methods. These methods may also be used for the evaluation of the quality of design and the construction of web pages.

This approach is supported by Nielsen (1993:223) who states that "there are many projects that would benefit from employing more than the minimum amount of discount usability methods" and "(usability methods) are intended to supplement each other, since they address different parts of the usability engineering life cycle, and since their advantages

and disadvantages can partly make up for each other. It is therefore recommended not to rely on a single usability method to the exclusion of others" (Nielsen, 1993:223).

It was therefore decided that the research problem would be addressed by applying multiple methods supplementing each other.

Other factors that influenced the choice of evaluation methods included:

- The number of users who were available for the evaluation – according to Nielsen "the choice of method may be dependent on the number of users available for usability activities" (Nielsen, 1993:225).

- The purpose of the evaluation and external limitations imposed on the evaluation process. The main goal was to collect, assess and report insightful feedback on what worked well and what did not in a relatively short period of time and to collect qualitative feedback. Therefore, discount usability methods, which allow for the gathering of feedback from small-sized samples, were included in the methodology. This was supplemented with a quantitative feedback method to collect feedback on respondents' expectations and perceptions about the site.

- The stage in the development process and the time and resources available (Bevan et al, 1991:2).

The evaluation of the *SA Government Online* website was thus done by applying the following evaluation methods:
- heuristic evaluation, including evaluation of the site by experts, and a critical evaluation of the website by the author against the set criteria
- user testing
- survey by means of an online questionnaire.

19

4.   DEVELOPMENT OF TEST INSTRUMENTS

4.1   Heuristic evaluation – expert opinion

4.1.1   Introduction

Nielsen (1993:155; 1994a) uses the term 'heuristic evaluation', which he defines as "a systematic inspection of a user interface design for usability" with the goal to "find the usability problems in a user interface design so that they can be attended to as part of an iterative design process." According to him heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognised usability principles, or the 'heuristics'.

Preece describes expert opinion as a diagnostic model lying between the theoretical approach taken in analytic evaluation and more empirical methods such as observational and experimental evaluation. She defines 'experts' as "people experienced in interface design or human factors or both" (Preece, 1993:111). The first part of the heuristic evaluation was the evaluation of the website by four 'expert' evaluators during the period 23 August to 8 September 2000.

4.1.2   Selecting evaluators

Four evaluators were chosen by using the following principles provided by Preece (1993:111) as guidelines:

- To ensure an impartial opinion the experts should not have been involved in the development of the system under evaluation.
- The experts should have suitable experience.
- The tasks undertaken by the experts should be representative of those intended for real users.

4.1.3   Evaluation guide

A semi-structured evaluation guide, including broad usability principles, was compiled and provided online. It was based on the criteria list referred to in paragraph two and included the five broad usability criteria, each with a few indicators to define it. Preece (1993:112) refers to such a guide as "predefined categorisation" where "experts are given a list of

problem categories". This approach is consistent with Nielsen's (1993:19) view. For discount usability testing, he advocates cutting the collection of usability guidelines to a small set of broader heuristics or basic usability principles. The evaluation guide also provided information on the aim of the evaluation, the aim of *SA Government Online* and instructions on how to perform the evaluation. The intention with choosing this form of reporting was to provide a platform that invited spontaneous comments and suggestions from the evaluators. The evaluation guide is attached as Annexure A.

4.1.4 Implementation

The evaluation was conducted during the period 23 August to 8 September 2000. The evaluators were asked to use the website for approximately three to four hours. They were provided with an online evaluation guide and asked to evaluate the website according to the usability principles as presented in the guide. They were also asked to consider any additional usability principles that came to mind and to describe potential problems they envisaged inexperienced and less experienced users might encounter while using the website.

4.1.5 Limitations and advantages of this method

According to Preece (1993:112), it is important to be aware of the following potential drawbacks of expert evaluation:

- Experts are often renowned for their strong views and preferences, in other words, biases.
- It is often difficult to find a person experienced in both a particular type of application and human computer interaction (HCI) research.
- Expert evaluation cannot capture the variety of real users' behaviour, whereas novice users can do unexpected things.
- Good role-playing requires an extraordinary amount of information about the knowledge level of the users, their typical tasks and their responses to problems.

However, this evaluation method was chosen because it would provide qualitative answers about the interface and would give a high potential return. "Even by using a small number of experts, they can usually identify a whole range of potential problems for users during a single session with an interface" (Preece, 1993:111). Additionally, expert evaluation is usually less costly than methods that involve user testing, and experts often suggest solutions to the problems that they identify (Preece, 1993:112).

4.2 Heuristic evaluation – critical evaluation of the website

The second part of the heuristic evaluation was a critical evaluation of the website by the author. The list of criteria and indicators was used for this method, and the content and features of the website were compared against the criteria with the aim to determine to what extent the website complied with them. The evaluation was conducted in the period 23 August to 8 September 2000.

4.3 User testing

4.3.1 Introduction

According to Preece (1993:112) expert evaluation cannot capture the variety of real users' behaviour, while Nielsen (1999h) states that there is only one good method to gather usability data, and that is to observe real users as they use the site to accomplish real tasks. Furthermore, according to Nielsen (1993:165), "user testing with real users is the most fundamental usability method and is in some sense irreplaceable, since it provides direct information about how people use computers and what their exact problems are with the concrete interface being tested" (Nielsen, 1993:165). Heuristic evaluation was thus complemented with the application of user tests.

This research made use of the discount usability engineering approach as described by Preece (1993), Nielsen (1993, 1994a) and Gordon (2000). This type of user test is done by involving real users, giving them typical work tasks and combining it with a verbal protocol where they are asked to think out loud while they perform the tasks. A verbal protocol is used to obtain a wide range of information, for example the user's planning for the particular task, what he/she is doing with the interface and why he/she is doing it, and to recall commands and arguments by the user during the test (Preece, 1993:114). This insight into a user's thought process might help pinpoint concrete interface elements that could cause misunderstandings.

The intention with user testing was to do a formative evaluation of the user interface in order to improve it. It therefore aimed to determine which aspects of the interface were good and which were bad and how the design could be improved (in contrast to summative evaluation which aims at assessing the overall quality of an interface). It is also worth noting that, because the intention was to collect qualitative feedback and not quantitative,

the objective was to do a task-based analysis, collecting feedback on how users tackled the tasks given, on where the major difficulties lay and on what could be done to improve the website (in contrast to a performance-based analysis which seeks to obtain clearly defined performance measures from the data collected).

4.3.2 Selection of respondents

With the discount usability engineering approach, a small number of respondents are tested. According to Nielsen, it is possible to apply the "thinking aloud" method effectively to evaluate user interfaces with a minimum of training and that "even fairly methodologically primitive experiments will succeed in finding many usability problems" (Nielsen, 1993:18). According to the Nielsen-Landauer cost-benefit model (Nielsen, 1994a), the maximum benefit-cost ratio is achieved when using between three and five users. Nielsen (1998a) states that a usability test with five users will typically uncover 80% of the site-level usability problems.

It was therefore decided to involve five respondents in this test. Respondents were selected from different user populations to cover different categories of users. Care was taken to ensure that they were representative of the most common user populations, i.e. average users and not users from outlier groups (Nielsen, 1993:175). The respondents were selected to have various levels of Internet and *SA Government Online* experience, and to be from different age groups and different genders. This provided a mix of individuals with varied skills and experience.

Respondents were chosen from persons who were easily accessible, yet representative of actual users of the website. The first respondent was an information centre worker in the GCIS who had high Internet and *SA Government Online* exposure. The second respondent was a journalist who had average exposure to *SA Government Online* and the Internet. The third respondent had been in the employ of the GCIS for less than a month and although he had high Internet exposure, his exposure to *SA Government Online* was very limited. The fourth respondent, representing the international user, had average exposure to the Internet as well as of the website. The fifth respondent was a lecturer at a university, who had high exposure to the Internet but none to the *SA Government Online* website.

4.3.3 Compilation of work tasks

Work tasks for this test were chosen to be as representative as possible of the actual use of the website and to provide reasonable coverage of the most important parts of the user interface. The work tasks were based on the intended uses of the website, usage statistics and observation of how users actually used the website. Work tasks were put in writing to ensure that all respondents received the tasks described in the same way, and to allow them to refer to the task description during the test (see Annexure B).

4.3.4 Pilot testing

To ensure that respondents, evaluators and users interpreted the questions or tasks correctly the user guide was subjected to pilot testing before the actual evaluation was done. Two pilot subjects were used, one with high exposure to the Internet as well as to *SA Government Online*, the other with average experience of the Internet and no experience of the website. After the test, the pilot subjects were debriefed on the comprehensibility of the test tasks and then requested to complete the online questionnaire (see paragraph 4.4). Thereafter the tasks were revised to fix difficulties found during the pilot activity and to ensure that the test instrument was easy to understand and to apply. One question was removed, the sequence of the questions was changed and certain questions were rephrased to make them more understandable.

4.3.5 Implementation

The user tests took place during the period 14 to 21 August 2000. Tests were conducted individually for each respondent in his/her normal work environment. Each respondent was tested for approximately 60 to 90 minutes. The respondents were briefed about the procedure before the evaluation commenced. They were given an overview of what a usability evaluation is, how it works and what they could do to help the evaluator to get the best information. Respondents were asked to be honest in their responses, not to do things any differently than they would if they were at home or at work, and ensured that it was not they who were being evaluated, but the website.

During the test, respondents were asked to give their initial impression of the home page and then given the opportunity to explore the website freely. The directed tasks (of which the first was very simple) that required of them to find some specific information on the website were then given to respondents. Respondents were allowed to ask questions about

the tasks in order to minimise the risk of misinterpretation. During the test respondents were encouraged to verbalise their thoughts (what they were doing and why) while performing the tasks. The experimenter did not interfere with the respondents, but allowed each to discover solutions on his/her own. However, the procedure was kept flexible to allow deviations from the questionnaire when respondents' actions determined the necessity thereof. When respondents supplied useful information that had not been anticipated in the guide, the particular route was explored further, and when they encountered problems, they were asked what they thought were ways to fix them. Body language and facial expressions were observed and followed up with questions when necessary.

Contrary to the traditional thinking aloud method where the procedure is videotaped the experimenter took notes: "In discount usability engineering we don't aim at perfection; we just want to find most of the usability problems" (Nielsen, 1993:18). Problems the respondent experienced, suggestions on how to fix them and specific quotes from the respondents were documented.

After completion of the tasks, respondents were asked to complete the satisfaction questionnaire (paragraph 4.4) and then debriefed. During the debriefing they were asked for comments about the website and suggestions on how to improve it. The experimenter also used the debriefing session to ask respondents for further comments about events during the test that the experimenter could not understand.

4.3.6 Choice of experimenter

The user test was conducted by the author. The advantage of using someone involved in the design of the website as experimenter was that she had knowledge of the interface. This enabled her to understand what the respondents were doing and to make useful inferences about the participants' probable intentions at various stages. An additional advantage was that "the experience of seeing users struggle with the system always has a very powerful impact on the designers" (Nielsen, 1993:180). A disadvantage of this method, however, might be a possible lack of objectivity that may cause an experimenter to help users too much.

25

4.3.7 Limitations and advantages of this method

In general, one of the limitations of applying user tests is that in-depth analysis can be very time-consuming. Notes have to be transcribed and data categorised. In addition, Macleod (1994:9) states: "While observation can yield rich usability data, analysis solely in real time may lose valuable data." Another limitation may be that user testing is an obtrusive method. Users are aware that their performance is being monitored and this can alter their performance levels (Preece, 1993:113). An important factor in the use of this method is thus the trade-off between time spent and depth of analysis (Preece, 1993:114).

Other methodological pitfalls are the issues of reliability and validity. Reliability is whether one would get the same result if the test were to be repeated and is a problem because of individual differences between test respondents. "One often needs to make decisions on the basis of fairly unreliable data, and one should certainly do so since *some* data is better than *no* data" (Nielsen, 1993:166). Validity is whether the result actually reflects the usability issues one wants to test. A high level of validity requires methodological understanding of the test method as well as some common sense (Nielsen, 1993:169). Typical validity problems involve using the wrong users or giving them the wrong tasks or not including time constraints and social influences.

In addition, thinking aloud may give a false impression of the cause of usability problems if too much weight is given to the user's own "theories" of what caused problems and what would help (Nielsen, 1993:195).

However, user-based evaluation can offer significant benefits in terms of the quality and validity of the usability data obtained (Macleod, 1994:9). It quickly highlights difficulties and provides qualitative data from a small number of users (Preece, 1993:119; Nielsen, 1993:195).

4.4 Online survey

4.4.1 Introduction

According to Macleod (1994:7), one of the simplest means of testing usability is to ask users, to sample their subjective views. This is confirmed by Nielsen (1993:209), who states that this is especially true for issues relating to users' subjective satisfaction and possible anxieties, which are hard to measure objectively. According to Macleod (1994:7),

this can be achieved in a structured way by using a questionnaire. Properly conducted and analysed, and with due consideration of contextual factors, valid and reliable measures of performance can be provided. This method is primarily a quantitative measurement of how usable a system is in the view of the user.

4.4.2 Compiling the questionnaire

An 18-item questionnaire was compiled and presented interactively on the website. Despite being mainly a quantitative measuring instrument, provision was made for qualitative questions as well. A combination of open and closed questions was included to allow respondents not only to select answers from a choice of alternative replies (quantitative), but also to provide them with the opportunity to give reasons for certain answers (qualitative). The questionnaire made use of the multi-point Likert rating scale to measure the strength of agreement against clear statements. Questions focused on information coverage, currency, ease of finding information on the website, 'look and feel' and preference with regard to search versus browsing techniques. The questionnaire was kept short and simple to promote the response rate and took between 10 and 15 minutes to complete. The questionnaire is attached as Annexure C.

4.4.3 Pilot testing

During an interview the interviewer can continuously evaluate the user's replies, making it possible to rephrase questions that seem to have been misunderstood. In contrast, questionnaires have to stand on their own. It is therefore essential that a questionnaire be subjected to pilot testing and iterative design before it is distributed, to ensure it is not too long, too hard to understand or too unprofessional (Nielsen, 1993:212). The questionnaire was thus subjected to pilot testing before it was implemented live on the website. It was digitised before pilot testing took place to ensure that the pilot test was conducted in a situation as close to real-life as possible.

The same two pilot subjects chosen for the user test were used. The pilot subjects were requested to complete the online questionnaire and submit it electronically. They were then debriefed on the comprehensibility of the questionnaire. After the pilot test the questionnaire was revised to correct difficulties experienced by the pilot subjects. The user feedback message and some questions were rephrased and two questions removed.

4.4.4 Implementation

The questionnaire was made live on the website on 14 August 2000 and was kept on the site until 16 September 2000, thus for a period of almost five weeks.

4.4.5 Limitations and advantages of this method

A possible limitation of this method could lie in the validity of the response rate. According to Crabtree (2000a), it is virtually impossible to construct a viable sampling frame of e-mail addresses. Even if a way to do it becomes available, users in lower socio-economic strata are not likely to have regular Internet access, and it will be some time before regular Internet access reaches an acceptable level of coverage. The degree to which results can be generalised to the larger population is thus a thorny question. It was also not possible to determine if a representative response rate had been achieved, as the size of the user population was not known. However, according to Crabtree (2000a), general representativeness should not be too big a concern when the survey is conducted among the segment of the population that used the Internet regularly.

Another important factor is that there is a difference between service quality and satisfaction. From a usability perspective a questionnaire is an indirect method since it does not study the user interface itself, but only users' opinion about the interface. One cannot always take user statements at face value (Nielsen, 1993:209). This evaluation method can therefore not be used to study the user interface, but only measures the perceived usability of the system. It will thus not provide detailed results with regard to usability and specific content issues.

Further limitations of this method could be the following:

- Privacy issues – concerns about security, both real and perceived, may provide obstacles. "Assurances of confidentiality ring somewhat hollow to many respondents in the largely unregulated online environment" (Crabtree, 2000a). More advanced online security techniques are helping to alleviate such fears.
- The possibility of a low response rate.
- Possible response bias.
- It might be difficult to analyse, even though open questions provide a rich source of data.
- Quantitative data do not necessarily give a valid indication of the level of usability (Macleod, 1994:3).

However, it should be noted that this data collection technique was chosen to address users' general satisfaction and attitude towards and understanding of the website. Despite its limitations, it can assist in determining the overall social and practical acceptability of the site amongst real users, as well as their attitude towards the aspects raised by the research questions. The use of results supplementary to those of the rest of the evaluation methods provided a complete picture of the most important issues to address.

5. SCHEDULING OF ACTIVITIES

To synchronise data gathering activities for the different methodologies all were scheduled to take place within the same period (14 August to 16 September 2000). Care was taken that data gathering did not take place over a period where heightened interest in the site could have influenced frequency of user visits or satisfaction with information on the site.

6. CONCLUSION

To evaluate the *SA Government Online* website, the author developed an evaluation methodology by following three main activities, i.e. the identification of evaluation criteria and guidelines, the selection of suitable evaluation methods, and the development of the test instruments for data collection.

The author developed a comprehensive list of criteria to be used for the development of the four test instruments used in the evaluation process. Criteria were developed that were applicable for the evaluation of any website, irrespective of which type of website it is, and also for the specific type of website, i.e. the gateway or entry point to other South African government websites. These criteria are presented in chapters four and five.

Although usability testing is a highly recommended evaluation method for websites, there are several other methods that could and should be used to gather supplementary data (Nielsen, 1993:207). To ensure as comprehensive as possible data gathering about the effectiveness and usability of *SA Government Online,* multiple methods were chosen for the evaluation of *SA Government Online.* The intention was that the methods should supplement each other, since they address different parts of the usability engineering lifecycle, and since their advantages and disadvantages partly make up for each other. The findings for each of the four evaluation methods are discussed in chapter six.