Measuring the quality of academic library electronic services and resources

Jillian R Griffiths

Research Associate CERLIM, Manchester Metropolitan University, Geoffrey Manton Building, Rosamond Street West, Manchester, M15 6LL, UK.
Email: j.r.griffiths@mmu.ac.uk, Tel: +44 (0)161 247 6142

Abstract

As a profession we are constantly striving to ensure that we provide the best services and resources possible to our users. This concern has resulted in a myriad of approaches and methods being utilised in an attempt to establish the quality of services and resources and lead to improvements in them. Online resources in particular have been the focus of much research in recent years, with work being undertaken in many areas including, for example, information retrieval, information seeking behaviour and usability studies, different approaches which share the same ultimate goal of making resources and systems easier to use by end users.

As a result of the shift in recent years from the usage of performance indicators to measures of outcome and impact within libraries (Brophy, 2004), a Quality Attributes approach is proposed in this paper. This approach allows for a holistic assessment of the quality of services or resources and encompasses usability. The classic definition of quality as "fitness for a purpose" was developed by Garvin (1987) into an eight dimension, or attribute, model, which can be used as a framework for determining the overall quality of a product or service. This approach has since been adapted for use in libraries and information services by Marchand (1990), Brophy and Coulling (1996), Brophy (1998), and Griffiths and Brophy (2002, 2005). Griffiths and Brophy adapted the Quality Attributes further by changing the emphasis of one attribute, changing the concept of one attribute, and introducing two additional attributes (Currency and Usability), thus producing a set of ten attributes which can be used to assess the quality, usability and impact of services and resources. These attributes are: Performance, Conformance, Features, Reliability, Durability, Currency, Serviceability, Aesthetics, Perceived Quality, and Usability.

This approach has been used to assess existing, and developing academic services of the UK Joint Information Systems Committee (JISC) Information Environment, results of which

can be seen in Griffiths and Brophy (2005, 2002). The paper will discuss use of these attributes within the context of evaluation of academic library services and resources.

Thematic section
Studies, Researches and Theoretical Approaches in the Evaluation of Academic Libraries: Researches of Satisfaction of Users, Quantitative and Qualitative Methodologies of assessment of sufficiency of Collections, Infrastructures, Electronic Services k.a.

Introduction
Walk into any lecture theatre in a UK higher education institution and ask what students use to find information and the majority will shout 'Google' – a situation we refer to as the 'Googling phenomenon' (Griffiths and Brophy, 2005). Students' use of search engines influences their perception and expectations of all electronic resources but, whilst the preference for very simple search engine approaches is prevalent, it does not mean that students are necessarily best served by this approach. Exclusive use of any commercial search engine, coupled with a lack of awareness and understanding of peer-reviewed, quality resources is not in the best interest of students or academic staff.

This issue of 'quality' plays a key role in professional discussions regarding electronic (and other) services and resources and as a result sophisticated approaches to quality assurance have been developed, such as peer review (largely for journal articles) and publisher reputation (for monographs) and relating these to the known interests of the user community, present and future. As a profession we are constantly striving to ensure that we provide the best services and resources possible to our users and this concern has resulted in a myriad of approaches and methods being utilized in an attempt to establish the quality of services and resources and lead to improvements in them. Online resources in particular have been the focus of much research in recent years, with work being undertaken in many areas including, for example, information retrieval, information seeking behavior and usability studies.

Evaluation of comparative online systems, services and resources has a long tradition of improving the state of the art in information retrieval (IR) technology. The criterion for the evaluation of performance effectiveness has largely been based on the overall goal of a retrieval system, that is, the retrieval of relevant documents and suppression of non-relevant items. Such evaluations adopt the Cranfield experimental model based on relevance, a value judgment on retrieved items, to calculate recall and precision. These dual measures are then presented together where recall is a measure of effectiveness in retrieving all the sought information and precision assesses the accuracy of the search.

Many and varied criticisms and concerns have been leveled at the validity and reliability of a Cranfield approach to IR evaluation (for example, Ellis, 1984). The core concern centers on the compromise necessary in the definition of relevance for such experimentation. That is, it is necessary to assume that relevance judgments can be made independently and that a user will

assess each document without being affected by their understanding of any previous items read. In addition, there is a basic assumption that relevance can ignore the many situational and psychological variables that in the real world affect it (Large *et al.,* 1999). The appropriateness of this traditional model has also been questioned when used in the Internet environment, the major limitation being that the Internet does not provide for a controlled environment. As a result many studies in this area use the precision measure only (Leighton and Srivastava, 1999; Clarke and Willett, 1997).

The Quality Attributes approach allows for a holistic assessment of the quality of services or resources and encompasses usability and evaluation of a user centered view of performance effectiveness using the user's own perception of relevance and satisfaction with both items retrieved and of the resource as a whole. The Quality Attributes approach has been adapted for use in libraries and information services by a number of authors, that is, Marchand (1990), Brophy and Coulling (1996), Brophy (1998), and Griffiths and Brophy (2002, 2005). The aim of this paper is to reflect on the work that has been undertaken in this field and to offer some observations on possible future directions. It explores the Quality Attributes approach pioneered by Garvin and presents the development of this approach from its original conception by Garvin to its latest adaptation by Griffiths and Brophy (2002, 2005).

Garvin's original Quality Attributes
In the general management literature, the classic definitions of quality emerged from the work of a number of researchers and writers, known collectively as the quality gurus (see Brophy and Coulling, 1996, Chapter 2). They developed a series of statements about and definitions of the concept of quality, of which the most commonly quoted are 'quality is fitness for purpose' and 'quality is conformance to requirements'.

Interest in quality management among librarians reached its peak in the 1990s. Researchers and practitioners across a number of countries undertook research in relation to the more general quality management approaches and practices to libraries. This interest was epitomised by a Total Quality Management conference held as part of the Library Technology Fair at Hatfield, U.K., in 1993. Other examples of work in the field include the UK Institute of Management guide to ISO 9000, the international quality management standard (Ellis and Norton, 1994) and Hernon and Altman's (1998) exploration of the application of quality management approaches.

The classic definitions of quality, concerned with 'fitness for the user's purpose', clearly relate to the benefits that customers perceive as arising from a service or product which they purchase or use and as such is more concerned with measures of impact and outcome rather than performance measures or indicators. Garvin recognised that the classic definitions of quality were too simplistic and set about developing a more sophisticated approach in the USA in the 1980s. He recognised that quality is a complex and multifaceted concept and suggested that there are eight critical dimensions or attributes that can be used as a framework for determining the overall quality of a product or service (Garvin, 1987).

Garvin suggested that many of the problems of defining and recognising quality arise because the concept can be approached from many different perspectives. He suggested that at least five views can be identified in the literature and in practice:

- the transcendental view: quality can be recognised, but cannot be defined;
- the customer view: quality as fitness for the customer's purposes or conformance to the customer's requirements;
- the manufacturer view: quality as conformance to specification;
- the product view: quality is related to inherent characteristics of the product;
- the value-based view: quality is dependent on what a customer is willing to pay for it. (Garvin, 1984)

Quality Attributes adapted

While Garvin's quality attributes approach was originally intended mainly for manufacturing industries it has subsequently been adapted for use in libraries and information services by Marchand (1990) and by Brophy (1998); Griffiths and Brophy (2002) and Griffiths (2003). The latter team suggested that library and information services might be assessed on the basis of ten quality attributes. Other work has suggested that this approach is also applicable to web pages (for example, Abels, White and Hahn, 1997; Madu and Madu, 2002).

Table 1.1: Garvin's original attributes contrasted with adaptations by Brophy and Griffiths

GARVIN	BROPHY/GRIFFITHS
Performance, the primary purpose of the product or service and how well it is achieving that primary purpose.	Performance, concerned with establishing confirmation that a library service meets its most basic purpose, such as making key information sources available on demand.
Features, secondary characteristics which add to the service or product without being of its essence.	Features: aspects of the service which appeal to users but are beyond the essential core performance attributes, such as alerting services.
Reliability, the consistency of the product or service's performance in use.	Reliability, which for information services would include availability of the service. Such problems as broken Web links, lack of reliability and slowness in speed of response would be measured as part of this attribute.
Conformance, whether or not the product or service meets the agreed standard, which may be internally or externally generated.	Conformance: whether the service meets the agreed standard, including conformance questions around the utilisation of standards and protocols such as XML, RDF, Dublin Core, OAI, Z39.50 etc.
Durability, the amount of use the product or service can provide before it deteriorates to a point where it needs replacement.	Durability, related to the sustainability of the information or library service over a period of time.
	Currency of information, that is, how up to date the information provided is when it is retrieved.

GARVIN	BROPHY/GRIFFITHS
Serviceability, how easy it is to repair a product or correct a service when it goes wrong, including the level of inconvenience experienced by the customer.	Serviceability, which may translate to the level of help available to users during, for example, information retrieval, or otherwise at the point of need. The availability of instructions and prompts throughout an online service, context sensitive help and the usefulness of that help could be measured in order to assess performance under this attribute.
Aesthetics, the appearance of the product or service.	Aesthetics and Image, both of the physical library and of web-based services based upon it.
Perceived quality, in essence the reputation of the product or service among the population, especially those with whom the potential customer comes into contact.	Perceived Quality: the user's view of the service as a whole and the information retrieved from it. It may be useful to measure perceptions both before and after a service is used.
	Usability, which is particularly relevant to electronic services and includes issues of accessibility.

Griffiths and Brophy adapted the Quality Attributes further by changing the emphasis of one attribute, changing the concept of one attribute, and introducing two additional attributes (Currency and Usability), thus producing a set of ten attributes which can be used to assess the quality, usability and impact of services and resources. These attributes are: Performance, Conformance, Features, Reliability, Durability, Currency, Serviceability, Aesthetics, Perceived Quality, and Usability. Usability, often used as an assessment criteria in its own right, has been defined by ISO 9241-11 as 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use' and, as Nielson points out, 'It is important to realize that usability is not a single, one-dimensional property of a user interface. Usability has multiple components and is traditionally associated with these five usability attributes: learnability, efficiency, memorability, errors, satisfaction' (1993, p.26). The following describes in detail the quality attributes as adapted by Brophy and Griffiths in the context of measuring the quality of academic library electronic services and resources.

Performance is concerned with establishing confirmation that a service meets its most basic requirement. These are the primary operating features of the product or service. For example, an electronic information service would be expected to retrieve a set of documents that matched a user's query. The most basic quality question is then 'Does the service retrieve a list of relevant documents?'

With **Conformance** the question is whether the product or service meets the agreed standard. This may be a national or international standard or locally determined service standard. The standards themselves, however they are devised, must of course relate to customer requirements. For information services there are obvious conformance questions around the utilisation of standards and protocols such as XML, RDF, Dublin Core, OAI, Z39.50 etc. Many conformance questions can only be answered by expert analysts since users are unlikely

to have either the expertise or the access needed to make technical or service-wide assessments.

Features are the secondary operating attributes, which add to a product or service in the user's eyes but are not essential to it. They may provide an essential marketing edge. It is not always easy to distinguish Performance characteristics from Features, especially as what is essential to one customer may be an optional extra to another, and there is a tendency for Features to become Performance attributes over time – inclusion of images into full text databases is an example of a feature developing in this way.

Users place high value on the **Reliability** of a product or service. For products this usually means that they perform as expected (or better). For electronic information services a major issue is usually availability of the service. Therefore broken links, unreliability and slowness in speed of response (speed of response of the service, a measure which has previously been reported as being important to users by Dong and Su (1997) can have a detrimental affect on a user's perception of a service.

Garvin uses the term **Durability**, defined as 'the amount of use the product will provide before it deteriorates to the point where replacement or discard is preferable to repair'. In the case of electronic information services this will relate to the sustainability of the service over a period of time. In simple terms, will the service still be in existence in three or five years? This is more likely to be assessed by experts in the field than by end users (although they may have useful contributions on the assessment of the attribute based on comparisons with similar services).
For most users of electronic information services an important issue is the **Currency** of information, that is, how up to date the information provided is when it is retrieved.

Serviceability relates to when things go wrong and is concerned with questions such as 'How easy will it then be to put things right', 'How quickly can they be repaired?', 'How much inconvenience will be caused to the user, and how much cost?' For users of an electronic information service this may translate to the level of help available to them during the search and at the point of need. The availability of instructions and prompts throughout, context sensitive help and usefulness of help were measured in order to assess responses to this attribute.

Whilst **Aesthetics and Image** is a highly subjective area, it is of prime importance to users. In electronic environments it brings in the whole debate about what constitutes good design. In a web environment the design of the home page may be the basis for user selection of services and this may have little to do with actual functionality.

Perceived Quality is one of the most interesting of attributes because it recognises that all users make their judgments on incomplete information. They do not carry out detailed surveys of hit rates or examine the rival systems' performance in retrieving a systematic sample of records. Most users do not read the service's mission statement or service standards and do their best to by-pass the instructions pages. Yet, users will quickly come to a judgment about the service based on the reputation of the service among their colleagues and acquaintances, their preconceptions and their instant reactions to it. Perceived Quality in the studies

undertaken by Griffiths and Brophy (2002) related to the user's view of the service as a whole and the information retrieved from it. This may be measured twice, before using the service (pre-perceived quality, where participants were aware of the service prior to testing) and after using the service (post-perceived quality). This allows investigation of how a user's perception of a service changes pre- and post-use.

The notion of quality is a concept which many students seem to struggle with. In the work undertaken for the EDNER study (Griffiths and Brophy,2002) it was found that 74% of students **did not agree** that the criteria '**Refereed'** was an indicator of quality. 89% agreed that 'Accurate' was an indicator of quality, 81% agreed that 'Current' was an indicator of quality and 52% that 'Reliable' was an indicator of quality. Students were also asked what quality meant to them and criteria such as Speed, Relevancy, Useful, Links to related areas, Understanding language used, Valuable, Referenced, Presentation, Timeliness, Accessible, Clear information and Source were listed (Brophy et al., 2003). Clearly there is confusion here.

The addition of **Usability** as an attribute is important in any user-centred evaluation. User-centred models are much more helpful when personal preferences and requirements are factored in and as such participants were asked how user friendly the service was, how easy it was to remember what the features/commands meant and how to use them, how satisfied they were with the input query facility and how satisfied they were with how to modify their query.

Quality Attribute measurement in practice - the process of assessment
Initial steps can be critical in the success of any assessment of services and preparatory work should not be underestimated. And, of course, the first place to start is identification of the resource or resources to be assessed. Deciding if a single resource or several resources are the focus will impact on:

- **why you are assessing** – are you concerned with how a single resource is performing? Are you trying to identify areas/elements where a single resource could be improved? Are you comparing resources to identify which is most effective, or best used, in order to inform collection decisions?
- **how you assess** – if assessing multiple resources the task undertaken on each resource needs, as much as possible, to be comparable.
- **who will assess**:
  - end users – public, students, academic staff
  - expert users – colleagues, usability/accessibility experts, you!
- **how you handle the resultant data** – how you analyse and work with your results will depend on what your focus is. Thinking about the sort of data you require, be it quantitative or qualitative, is essential in ensuring that the design of the assessment captures the data which will provide you with the information you require about the resources you are investigating.

Deciding on one-to-one testing or simultaneous multiple testing depends on a number of factors, including why you are assessing, what you are assessing and the resources and constraints within which you have to work. Single testing can yield greater qualitative data through observation and/or transaction logging (Griffiths et al. (2002) but it is likely that fewer

participants will be recruited due to the time and resource constraints. Simultaneous multiple testing is likely to yield more quantitative and less qualitative data but has the advantage of being able to engage with a greater number of participants, and it is always possible that some of these participants will agree to take part in focus group or other qualitative data collection techniques.

Design of tasks/test searches
In choosing the resource or resources it is most likely that consideration has been given as to why evaluation is being undertaken, and this in turn will affect the design of any tasks or test searches. As a general rule if assessment is being made to gain an understanding of users' behaviour then participants should be allowed to use their own tasks or queries. If the evaluation is to assess the service then it will be necessary to design tasks or test searches.

A task based approach can be very effective and can be 1) very directed, as in McGillis and Toms (2001) who used this approach to assess the usability of an academic library website by faculty staff and students or 2) looser simulations of real world situations such as those proposed by Borlund (2003) and developed from work by Ingwersen (1992, 1996) and Byström and Järvelin (1995). Here, a short 'cover story' describes a situation that requires the participant to use the electronic resource. The 'cover-story' is a rather open description of the context/scenario of a given work task situation. Where directed tasks are used care must be taken to ensure that they are of relevance to users, avoiding "tasks that focus almost exclusively on pet features rather than on goals that users really want to accomplish" (Nielsen, 2007). Tasks need to be piloted, timed, assessed and re-designed in an iterative cycle until you are satisfied that they test the service and all its aspects. They also need, as far as possible, to be of equal difficulty if assessing several resources.

Development of questionnaires for post-searching quantitative data collection
Questionnaires can provide large quantities of data, both quantitative and qualitative, in a relatively simple, structured and resource efficient way. Their advantages are numerous. They provide good structure; are less expensive than many other methods; are relatively easy to administer without special training; are not subject to interviewer bias; can produce quick results.
In addition online survey systems will undertake basic analysis. Their disadvantages include: it is more difficult than it appears to create a good, unambiguous questionnaire; response rates may be low - people often suffer from 'questionnaire fatigue'; there is no chance to ask a follow-up question if the initial answer is not sufficiently revealing.

Measuring the Quality Attributes
The following table presents examples of measures designed to provide responses to enable assessment of the service/resource for each Quality Attribute. These measurement instruments were developed to assess existing and developing services of the UK Joint Information Systems Committee (JISC) Information Environment as part of the EDNER and EDNER+ projects (Griffiths, and Brophy, 2005; Griffiths, 2003). This research utilized a user-centered approach, but could be tailored to different contexts and stakeholders, thus enabling a holistic approach to the evaluation of digital library services.
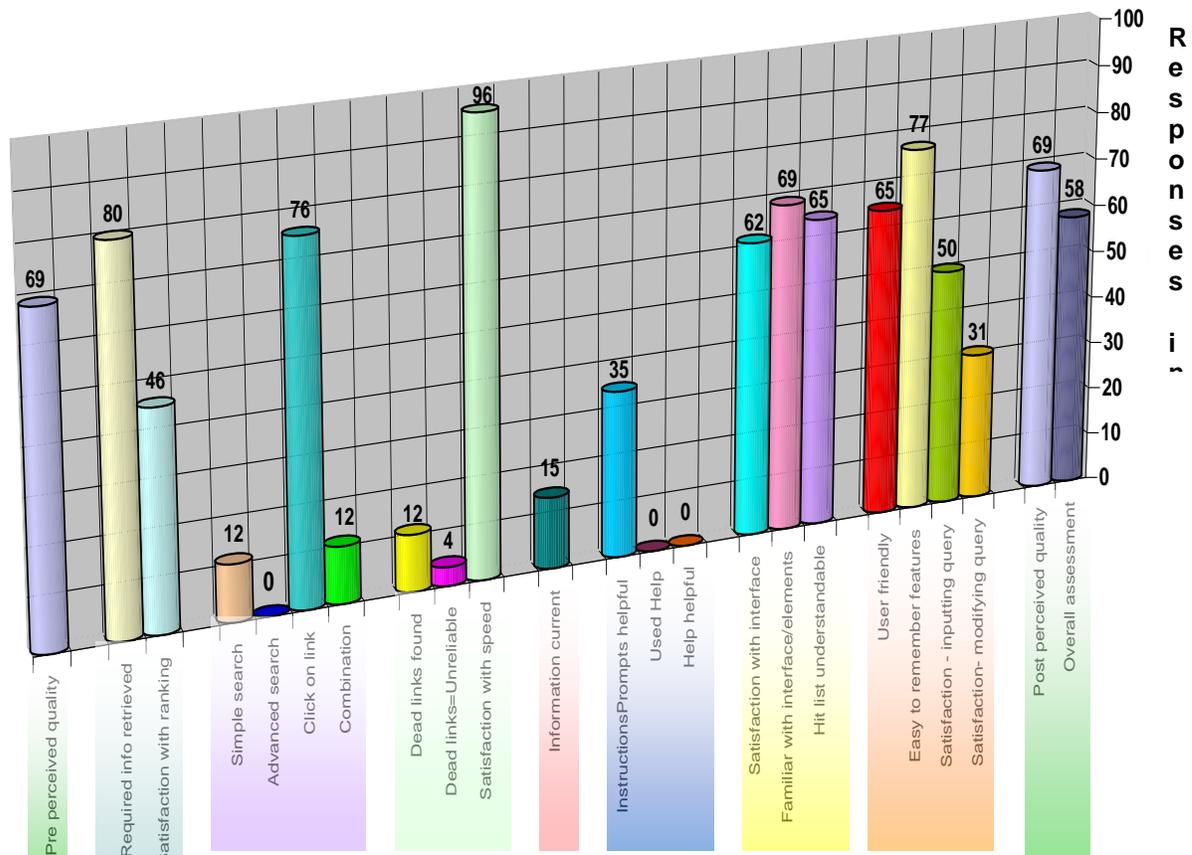
Table 1.2: Examples of Quality Attribute measurement instrument

Quality Attribute	Measurement
Performance	• Are you satisfied that you found the required information on this service? • How satisfied were you with the order of the items retrieved (e.g. most relevant first, least relevant last)?
Conformance	• Not evaluated by end users, could be assessed by expert user/service provider
Features	• There can be many different search options that you could have used to search for the information on this service. Which did you use (please list)? • Were there any features that you particularly liked about this service?
Reliability	• Did you find any dead links when using this service? • If Yes, does the finding of dead links in your search session make you think that the service is unreliable, reliable or doesn't it affect your judgment? • How satisfied were you with the speed of response of this service?
Durability	• Not evaluated by end users, could be assessed by expert user: assessment of long term service strategy
Currency	• Do you think that the information from this service is up-to-date?
Serviceability	• Did you find on-screen instructions & prompts helpful or unhelpful? • Did you select Help at any stage? • If Yes, how helpful did you find it?
Aesthetics	• How satisfied were you with the interface & features of this service? • Were there any features or aspects of the interface that you were familiar with? • How easy is it to understand what each item is about from the retrieved list?
Perceived quality	• Please rate the overall quality of the items you found on this service
Usability	• How user friendly did you think this service was? • How easy was it to remember which features to use? • How satisfied were you with the facility to input your query? • How satisfied were you with the facility to modify or change your query, e.g. find similar, related searches, refine search, etc?

Benefits of this approach

Assessment of a single service

One of the major benefits of the Quality Atrributes approach is that they allow investigation of how a user's perception of a resource changes pre- and post-use, and show that whilst preconceived notions of a service may be negative it is possible to change these perceptions if the resource performs well across a number of the Attributes. If pre-use perceptions do not alter it is possible to identify which aspects of a resource need to be improved by examining those Attributes that users have scored lower. For example, in Figure 1.1 a participant has used a service and then indicated their satisfaction across a range of Attributes, thus:

Figure 1.1 Example of a participant's assessment of resource

For this resource participants indicated high levels of satisfaction that the required information was retrieved (80%) but satisfaction with ranking was much lower (46%). The majority of users navigated to information by clicking on links (76%), with only 12% using the simple search option and 12% using a combination of techniques. A small number of participants (12%) found dead links and of these 4% felt the service was unreliable as a result. Satisfaction with speed of response was very high at 96%. Low levels of satisfaction were recorded on the Currency attribute (35%). Instructions and prompts were found to be helpful by 35% of participants and the Help facility was not used in any of the cases. 69% of participants reported that they were familiar with the interface, or elements of the interface and 62% of participants were satisfied with it. 65% felt that the hit list was understandable. On the Usability attribute 65% of students felt that the service was user friendly and 77% found features and commands easy to remember and use. 50% of participants reported that they were satisfied with the facility to input their query but only 31% were satisfied with the facility to modify their query. The preferred method of obtaining information was by navigating via click on link as opposed to engaging with a search option. Perceived Quality, pre- and post-searching, remained static at 69% and overall satisfaction was recorded at 58%.

These results show that measures other than just a service's Performance play an critical role in a student's perception of a service and resource developers, providers and educators need to

understand a user's reaction to individual components of a resource in order to develop and improve services which will meet users' expectations.

Assessment of multiple services

Assessment does not just have to be conducted on an individual resource, data gathered from assessments of a number of resources and services can assist in the identification of those which are underperforming or are being under used. In figure 1.2 (looking just at Performance, Usability, Aesthetics and Perceived Quality) it is possible to make comparisons which can inform collection development and retention.
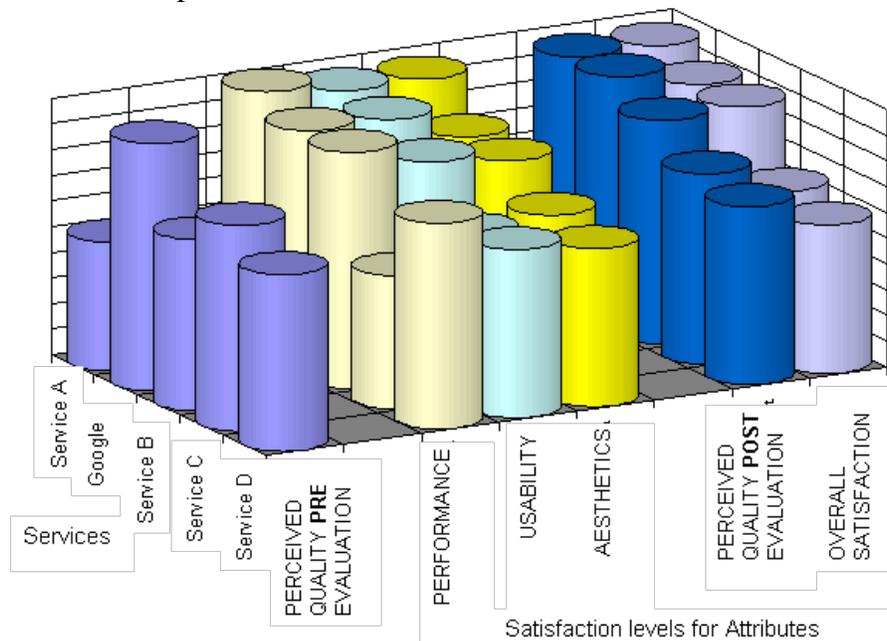

Figure 1.2   Comparative assessment of resources

Results show varying degrees of satisfaction across each of the services and each of the Attributes. On Service D, Perceived quality pre and post searching remained static (and relatively low), despite high levels of satisfaction with Performance. In conjunction with this Performance score, satisfaction with Usability and Aesthetics were lower. Overall Satisfaction was also relatively low. This seems to indicate that users' perceptions of quality are driven by factors other than Performance of a system. It also raises interesting questions as to how fixed preconceptions about quality may affect the results of the evaluation of a system or service.

On the University OPAC post-search Perceived Quality dropped only very slightly (2%), despite low levels of satisfaction with performance. Satisfaction with Usability and Aesthetics was slightly higher than that of Performance. This again may indicate that factors (or Quality Attributes) such as User Friendliness and Interface Design may be at least as important as evaluation criteria as Performance. Users expressed an increase in post-search Perceived Quality on Services A and C, coupled with high levels of Satisfaction Overall and across each of the Attributes. In this instance actual use of the service appears to have changed the preconceptions of the users.

Results for Google were high across all the measures used in the test. However, satisfaction with Aesthetics was lower than satisfaction with Aesthetics on Service C, and had lower Overall satisfaction ratings.

It is interesting to note that in each instance Satisfaction Overall corresponded very closely with post Perceived Quality

Conclusion

Previous research (Griffiths and Brophy, 2005) has indicated that students are confused as to the meaning of quality when it comes to assessing academic resources. Viewed in the light of the findings of Cmor and Lippold (2001), who stated that students will give the same academic weight to discussion list comments as peer-reviewed journal articles, it would seem that students are poor evaluators of the quality of academic online resources. The original premise of the Perceived Quality attribute is that users make their judgments about a service based on incomplete information and that they will come to this judgment based on its reputation among their colleagues and acquaintances and their preconceptions and instant reactions to it. If the notion of quality conveys so many different meanings to students, it poses something of a challenge to the academic community in encouraging students to understand and use quality-assured electronic resources.

It is also apparent, from a methodological perspective, that further work is needed to explore the meaning of Perceived Quality and the interpretation of user responses to this area of enquiry. Fundamentally different understandings of information quality could otherwise lead to questionable conclusions being drawn by researchers and service providers. Measures of Impact should also be taken into account and future work is planned to include this as an eleventh attribute. And so the Quality Attributes approach to assessment of library services and resources continues to evolve and change.

The assessment of service and resource quality has developed significantly in recent years. We have robust sets of performance indicators which provide the basic 'picture' of library performance and, beyond that, we have ways to explore the customer experience, to find out how users feel about the services and resources they use and to use these insights to provide better managed services. Brophy (2004) has ruminated that, as a profession, we may be moving beyond individual techniques in an attempt to synthesise the different approaches towards measurements of impact, to get back to the essential question of 'do libraries and their services do any good?', and, as Brophy says, our challenge is to strive to seek, and understand, the evidence that they do.

References

Abels, E.G., White, M.D. and Hahn, K. (1997) "Identifying user-based criteria for Web pages." *Internet Research: Electronic Networking Applications and Policy*, 7(4),pp.252-62.

Brophy, P. (1998) "It may be electronic but is it any good? Measuring the performance of electronic services." *Robots to Knowbots: the wider automation agenda. Proceedings of the Victorian Association for Library Automation 9th Biennial Conference, January 28-30 1998.* Melbourne, Australia: VALA, 1998, pp.217-230. http://www.vala.org.au/valaweb/num511.pdf

Brophy, P. (2004). The quality of libraries. In: Hilgermann, K. and te Borkhorst, P. (eds.) Die effective biobliothek: Roswitha Poll zum 65 Geburtstag, pp.30-46. Munich: Saur.

Brophy, P. and Coulling, K. R. (1996) *Quality Management for Information and Library Managers*, Aldershot: Gower.

Borlund, Pia (2003) The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, **8**(3), paper no. 152. http://informationr.net/ir/8-3/paper152.html

Brophy P, Fisher S, Griffiths J. R. and Markland M. (2003) *EDNER: Formative Evaluation of the Distributed National Electronic Resource: DNER Service evaluation* (Deliverable MDA 2, EDNER Project). Manchester: CERLIM. [Online] http://www.cerlim.ac.uk/edner/dissem/dissem.html

Byström, K. & Järvelin, K. (1995) Task complexity affects information seeking and use. *Information Processing & Management*, **31**(2), 191-213.

Clarke, S.J. and Willett, P. (1997). Estimating the recall performance of Web search engines. *Aslib Proceedings*, 49 (7), 184-189.

Cmor, D., & Lippold, K. (2001). *Surfing vs. searching: The Web as a research tool.* Paper presented at the 21st Annual Conference of the Society for Teaching and Learning in Higher Education, June 14–16, Newfoundland. Retrieved November 16, 2004, from http://staff.library.mun.ca/~DCmor/stlhe/.

Craven, J. and Griffiths, J.R. (2008). Involving users in a technical solution to help assess the accessibility of websites. In Brophy, P. et al. (eds.) *Libraries without walls 7: Exploring 'anywhere, anytime' delivery of library services*. (2008). Facet Publishing.

Dong, X. and Su, L. (1997). Search engines on the World Wide Web and information retrieval from the Internet: A review and evaluation. Online & CDROM Review, 21(2), 67-81.

Ellis, D. (1984).The effectiveness of information retrieval systems: the need for improved explanatory frameworks. *Social Science Information Studies*, **4**, 261-272.

Ellis, D. and Norton, B. (1993) *Implementing BS5750 / ISO 9000 in libraries.* London: Aslib.

Garvin, D. A. (1984) "What Does "Product Quality" Really Mean?" *Sloan Management Review,* pp.25-45.

Garvin, D. A. (1987) "Competing on the Eight Dimensions of Quality." *Harvard Business Review*, November-December, pp.108-109.

Griffiths, J. R., Hartley, R.J., Willson, J. P. (2002) An improved method of studying user-system interaction by combining transaction log analysis and protocol analysis. *Information Research*, **7**(4) http://InformationR.net/ir/7-4/paper139.html

Griffiths, J.R. (2003) Evaluation of the JISC Information Environment: student perceptions of services. *Information Research*, **8** (4). [Online] http://informationr.net/ir/8-4/paper160.html

Griffiths, J.R. and Brophy, P. (2002) "Student searching behaviour in the JISC Information Environment." *Ariadne*, 33, [Online] http://www.ariadne.ac.uk/issue33/edner/intro.html

Griffiths, J.R. and Brophy, P. (2005). Student searching behaviour and the web: use of academic resources and Google. *Library Trends*, Spring, pp. 539-554.

Hernon, P. and Altman, E. (1998) *Assessing service quality: satisfying the expectations of library customers*. Chicago: American Library Association.

Ingwersen, P. (1992) *Information retrieval interaction*. Taylor Graham.

Ingwersen, P. (1996) „Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, **52** (1), 3-50.

Large, A., Tedd, L.A. and Hartley, R.J. (1999). *Information seeking in the online age: principles and practice*. London: Bowker Saur.

Leighton, H.V. and Srivastava, J. (1999). First 20 precision among world wide Web search services (search engines). *Journal of the American Society for Information Science*, 50 (10), 870-881.

Marchand, D. (1990) Managing information quality. In Information quality: definitions and dimensions (ed. I. Wormell) London: Taylor Graham, pp.7-17.

McGillis, L. and Toms, E. G. (2001). Usability of the academic library web site: implications for design. *College and Research Libraries*. **62**(4), pp.355–367.

Madu, C.N. and Madu, A.A. (2002), "Dimensions of equality", *International Journal of Quality & Reliability Management*, 19 (3), pp.246-259.

Nielsen, J. (1993) *Usability Engineering*. Academic Press.

Nielsen, J. (2007). Should Designers and Developers Do Usability? *Alertbox*, June 25 www.useit.com/alertbox/own-usability.html

Peters, T. (1993) The history and development of transaction log analysis. *Library Hi Tech*, **11**(2), 41-66.