



IFLA  
2005  
OSLO

## World Library and Information Congress: 71th IFLA General Conference and Council

### "Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

*Conference Programme:*

<http://www.ifla.org/IV/ifla71/Programme.htm>

June 07, 2005

**Code Number:**

**050-E**

**Meeting:**

**108 Education and Training with Library Theory and Research**

#### **Division VII Education and Research**

### **The basis for evidence-based practice: evaluating the research evidence**

#### **Dr L. Anne Clyde**

Professor, Faculty of Social Science

The University of Iceland

101 Reykjavík, Iceland

Web: <http://www.hi.is/~anne/>

Email: [anne@hi.is](mailto:anne@hi.is)

***Abstract:** Evidence-based practice is one way in which research can influence professional practice. However, if practitioners are to use research evidence as a basis for practice, then they need to be able to evaluate the research evidence; that is, they need to be able to assess research quality. This paper describes the criteria, strategies and models that are available for assessing research quality, and discusses the limitations and problems associated with those strategies and models. A pilot study that investigated the ways in which experienced research evaluators assess research quality, is described, and its implications are discussed. The paper concludes with a brief description of the ongoing research that is building on the results of the pilot study.*

#### **INTRODUCTION**

Library and information science research intersects with professional practice very explicitly in the form of evidence-based practice. This paper provides an

international research perspective on evidence-based practice, focusing in particular on how the quality of the research evidence can be assessed as a basis for evidence-based practice. Evidence-based practice in librarianship emerged in the second half of the 1990s (Eldredge, 1997) as a way of improving professional practice and highlighting service outcomes that matter in the community served by the library. It focuses on using documented evidence from professional practice and from research to resolve day-to-day problems and to plan for the future. This then raises a number of questions in relation to the evidence, including the research evidence. Will any evidence do? Is some evidence better than other evidence? What is “good evidence”? And not least, how can a busy practitioner distinguish between quality research evidence and evidence that might be unhelpful or even false?

While many commentators have discussed research quality in library and information science (for example, McClure & Bishop, 1989; Herson, 1999; Haycock, 1994), there is nevertheless little agreement about what constitutes “quality” in a research publication in this field, nor how “quality” can be recognized or measured. Some exponents of evidence-based practice avoid the issue altogether by assuming that all published evidence will be reliable and useful, since it will in many cases at least have gone through a number of editorial processes before being published. However, this assumption becomes very difficult to sustain when obvious errors are found in published research reports (see, for example, Fazackerley, 2003), or when two published reports present apparently conflicting evidence, or when hoax articles are accepted for publication by supposedly reputable academic journals and conferences (see, for example, Sokal, 1996; Reuters, 2005). On the basis of a literature review and an ongoing research project (Clyde, 2004), this conference paper discusses current strategies and models for evaluating research publications, and the problems associated with those strategies.

## **EVALUATION OF PUBLISHED RESEARCH**

There are numerous lists of criteria for evaluating published research; however, they are generally descriptive rather than evaluative; that is, they describe features that an evaluator would expect to see in a research report, rather than providing indicators of quality. Thus, for example, according to Pálsdóttir *et al.* (1997, 88), a research report should include a statement of aims, research questions or hypotheses, a description of the methodology, information about data collection and analysis, discussion of the results, and a bibliography, among other things. However, while lists like this do help us to identify research reports, they do not take us very far down the road of evaluation. Even where evaluative elements are included, there is seldom any indication of how the evaluator might approach an assessment of each factor (see, for example, Day & Peter, 1994); Audunson (2004) is an exception in that he provides some examples from research projects to illustrate the application of the criteria.

More analytical approaches tend to rely on using the quality of the journal in which an article is published (and by analogy the quality of a research conference for which a paper is accepted) as a basis for research evaluation. The assumption behind these approaches is that “high-quality journals ... are likely to publish high-quality research” (Lee, *et al.*, 2002, 2805). Evaluation of research articles then becomes a matter of identifying the quality journals. Examples of strategies to do this include

citation analysis (see, for example, Garfield, 1979; Harter, 1996; Nicolaisen, 2002), journal impact factor analysis (see, for example, Opthof, 1997), the number and percentage of collaborative articles published in the journal — as in, articles written by more than one author (see, for example, Hart, undated), approaches based on the reputation of journals (see, for example, Giles, *et al.*, 1989; Blake, 1996; Kohl & Davis, 1985), peer-review status of the journal (Lee, *et al.*, 2002), manuscript acceptance rate (Lee, *et al.*, 2002), indexing of the journal in established indexing or abstracting services (Gehanno & Thirion, 2000), the number of subscribers to the journal (Lee, *et al.*, 2002), and the number of external links to the journal web site (Vaughan & Hysen, 2002).

All these strategies have their strengths and limitations. However, there are those who believe that while they have their uses, they should not be used for research evaluation. Moed (2002), for example, says, “bibliographic indicators reflect scientific impact, not quality”, though it could be said that they are not even very good at doing that (Moed, 2002). Gorman and Calvert (2000), reporting on factors that contribute to journal quality, say that “The fact that paper x is cited y times is not an indication of quality, but rather that it is cited — it is available, it is in a journal held by many libraries, the author (or publisher or editor) is particularly good at self-promotion”. MacRoberts and MacRoberts (1989) have studied “problems of citation analysis” in a critical review; Kostoff (1998) has discussed “the use and misuse of citation analysis in research evaluation”; while Warner (2000) suggested that the value of citation analysis lies only in its capacity to “inform, not determine, judgements of research quality”.

Meanwhile, in terms of the other strategies for research evaluation, Seglen (1997) has written an article titled “Why the impact factor of journals should not be used for research evaluation”. Bahr and Zemon (2000) have discussed problems associated with multiple authorships of articles as a way of indicating accountability for research results, while Avkiran (1997) has shown that “there is no significant difference between the quality of collaborative and individual research” (“collaborative” in this case meaning “more than one author or researcher” on the project). Lowi (1992) has criticized the “reputational” approach to journal evaluation on the grounds that the studies are based on subjective evaluations and perceptions of the participants; Nkereuwem (1997, p.75) noted that the “ranking of a journal will depend to a large extent on the values which one brings to the evaluation process”; it will also be influenced by the journals to which the participants have access. The type of peer review process employed by a journal has been shown to influence review quality and the type of articles that are accepted (Justice, *et al.*, 1998) though not necessarily research quality. Although some researchers have found a correlation between links to the web site of a journal and another indicator of journal quality, other researchers (for example, Thomas & Willet, 2000) have not. As indicators of research quality, manuscript acceptance rates, indexing and abstracting in key services, and number of subscribers to the journal, all have their limitations, though all provide useful information about the journals (for different purposes).

Alternatively, a research article will be assumed to be of good quality if it appears in a journal that is on a recognized or approved list, such as the international “ISI Master Journal List” (Institute of Scientific Information, 2004) or national lists such as the “Guide to Core Journals of China” (Calvert & Zengzhi, 2001). This is

despite the fact that it is the journals that are evaluated for inclusion in these lists rather than individual articles, and despite the fact that the criteria used for evaluating the journals tend to reflect the characteristics of the journals rather than of the articles they contain. An example of this is the ISI journal selection criteria:

“The evaluation process consists of evaluation of many criteria such as, Basic Journal Publishing Standards (including Timeliness of publication), adherence to International Editorial Conventions, English Language Bibliographic Information (including English article titles, keywords, author abstracts, and cited references.) ISI also examines the journal’s Editorial Content, the International Diversity of its authors and editors. Citation Analysis using ISI data is applied to determine the journal’s citation history and/or the citation history of its authors and editors.” (Institute of Scientific Information, 2002)

Katzer, Cook and Crouch caution against an approach to the evaluation of research quality that is based on evaluation of the journals. They say that consumers of research cannot “assume that an editor will successfully weed out all major errors and poorly constructed studies. The explosion of information, the reward structure in higher education, the increasing number of new journals being published, the evaluation procedures used by editors, and a variety of other factors” (Katzer, Cook & Crouch, 1998, 6) all mean that it is necessary that each published research report be evaluated on its own merits if the results of the research are to be used to improve practice or if the results of the assessment are to be used for purposes such as performance evaluation.

Some commentators avoid the issue of quality assessment altogether by assuming that all published evidence will be reliable and useful, since it will, in many cases at least, have gone through a number of review and editorial processes before being published. However, this assumption becomes very difficult to sustain when obvious errors are found in published research reports or when two research reports, published at about the same time, present apparently conflicting evidence. In a review article in *The Times Higher Education Supplement* (15 August 2003, 25), Henry McQuay, a Professor at Oxford, was blunt about modern research publishing; it has, he says, “no room for critical appraisal of evidence” and even “the most ludicrous ... findings” are given credence. Certainly in the last few years, we have seen published research reports — research reports that had been through a peer review process — subsequently publicly questioned or discredited. The topics of these research reports are very varied and include research into the MMR (measles, mumps, rubella) vaccine in the United Kingdom (Henderson, 2004, 2), hormone replacement therapy for women (Derbyshire, 2004, news9), prescribing anti-depressant drugs for children (Kirkey, 2004), the spread of pollen from genetically-modified maize in Mexico (Henderson, 2003, 10), and magazines for pre-teen girls (“ABC sore...”, 2004, 24). In a study of articles published in leading academic journals, carried out in the United Kingdom by Tooley and Darby (1998), “there was evidence of partisanship in the conduct, presentation and argument of a significant number of the articles considered” (Humes & Bryce, 2001, 335).

As a result, the peer review system on which the research journals rely to evaluate and select papers for publication and as “a safeguard against the communication of poorly-conducted research”, was called into question (Henderson, 2003, 10; Humes & Bryce, 2001, 335). Peters and Ceci (1982) challenged the

reliability of the peer review process more than 20 years ago, as a result of a study in which they resubmitted 12 articles to the journals which had originally published them (18 to 32 months previously) after changing the names of authors and institutions and other minor details. Of the 12 articles, only three were recognized. Of the remaining nine, eight were rejected (by the same journals that had originally published them). A research study by Wood and Roberts (2004), that investigated the peer review process for research conferences, found further disquieting evidence of problems. Sir Patrick Bateson, Vice President of the Royal Society is quoted as admitting that “peer review is not perfect, and some scientific papers [have been] approved by referees and published in leading journals but later shown to have been based on false or poorly interpreted results” (Henderson, 2003, 10); however, the emphasis in subsequent discussions has been on improving the peer review process as a guarantee of research quality rather than on developing an alternative. Perhaps it is time we did try to find alternatives. Though relatively cheap for the journals, the peer review process is expensive (in terms of their time) for the academics who act (for free) as referees, and for their institutions who support this appropriation of academic staff time which might be used for other purposes. While it is probably true that it is the failures of the peer review process that have captured the attention of the media in the last few years, nevertheless there have been enough well-publicized failures to suggest that the peer review process is not completely reliable as a guarantee of the quality of the research reports that make it through to publication.

## **MODELS FOR EVALUATING EVIDENCE**

Evidence-based practice is becoming accepted in fields as diverse as medicine, physical therapy, education, librarianship, and literacy development. In the different fields, evidence may be of different types; for example, in clinical medicine randomized trials and systematic reviews are considered “the gold standard” of evidence (Feinstein & Horwitz, 1997), while in school librarianship, evidence from documented professional practice is considered to be of value alongside evidence from published research studies (Todd, 2003). A number of models have been proposed for the evaluation of research in fields like medicine (for example, Bleakley & MacAuley, 2002; MacAuley, 1994) and education (for example, Viadero, 2002; Ross, 2002). Three of these models will be discussed briefly below, to indicate the difficulties associated with this approach as well as some potential benefits.

- **READER.** This model (MacAuley, 1994) comes from the field of clinical medicine, where it was developed to promote the adoption of an evidence-based medicine approach to medical practice. It relates specifically to the methodological concerns of research in clinical medicine, which emphasize, for example, randomized control trials, case-control, and cohort studies. It has been “used by doctors in a number of different environments” (MacAuley & McCrum, 1999) and validated through research (for example, MacAuley, McCrum & Brown, 1998). However, there are indications that READER provides guidance rather than a fail-safe strategy for evaluation. Lee *et al.* (2002) carried out a study in which they had participants assess the quality of 243 research articles chosen from medical journals (using the MEDLINE database); the READER strategies were used in the evaluations. The participants received training in using the READER forms and had access to detailed printed instructions as well. Two

participants looked at each article independently, and where the two scores for an article were significantly different, the participants were instructed to reach a single score by consensus. Overall, the level of agreement between the evaluators was only “fair”. And this despite the fact that articles that did not meet basic clinical research methodology requirements were excluded from the study on the grounds that READER was not designed to deal with them. Another experiment also showed that participants in a trial assessed two given papers very differently (MacAuley, McCrum & Brown, 1998). However, in this case, some 70 per cent of the 104 participants believed that READER would make them more critical readers of research in the future (MacAuley, McCrum & Brown, 1998). In fact, READER’s greatest value may be that it does seem to increase awareness of the need for critical assessment of research as a basis for professional practice.

- **CASP.** The Critical Appraisal Skills Programme (CASP) was developed in the Public Health Resource Unit of the National Health Service (NHS) in the United Kingdom. It aims to “enable individuals to develop the skills to find and make sense of research evidence, helping them to put knowledge into practice” (CASP, 2003). It promotes a three-step process of “finding research evidence, appraising research evidence, and acting on research evidence” (CASP, 2003). The CASP web site has a range of supporting resources and materials, but unfortunately not all links on the site were working at the time the analysis of models was carried out. According to the web site, CASP has separate evaluation forms for research articles based in different research methods — for example, systematic reviews, randomized controlled trials, qualitative research studies, cohort studies, case control studies, and diagnostic test studies. This means that evaluators have first to go through an article to determine which of the evaluation forms would be most appropriate, before the article could be evaluated. This, coupled with the detailed forms, means that the evaluation process is a lengthy one. However, like READER, it appears to have some value in raising awareness of the need for critical assessment of research as a basis for evidence-based practice.
- **CriSTAL.** “In 1999, staff at the universities of Sheffield [the School of Health and Related Research] and Oxford [the Health Care Libraries Unit] commenced an unfunded project to examine whether it is feasible to apply critical appraisal to daily library practice” (Booth & Brice, 2003, 45), and to enable librarians in the health care field to interpret and apply research as a basis for practice. Critical Skills Training in Appraisal for Librarians (CriSTAL) involves the application of the principles of evidence-based medicine to library practice (Booth, 2003). By the end of 2003, the CriSTAL project team had developed two checklists, both specific to particular kinds of research studies rather than methodologies. The first was related to user studies, the second to information needs analyses or information audits. These two were chosen for the initial work because of their importance in library and information science research generally. An evaluation, conducted within the context of a workshop for health librarians, showed that the checklists “helped participants improve their understanding of research methods and their ability to use research to aid their decision making” (Booth & Brice, 2003, 45). As was the case with CASP, the forms were lengthy and a pre-evaluation process was necessary in order to select the most appropriate checklist.

All three models have some value as tools for professional development, particularly in the context of developing research evaluation skills. It is no accident that all were trialed in workshop settings. However, while the models assist practitioners to develop skills in evaluation, the complex evaluation instruments and the need for pre-evaluation in two cases, make them cumbersome as tools for day-to-day decision-making. In addition, the research showing only “fair” levels of inter-rater agreement when the READER model is used, suggests that it provides no guarantees of research quality.

## **EVALUATING RESEARCH QUALITY: A RESEARCH PROJECT**

It seems that not only do people attempt to assess quality in different ways, but they also have different understandings of the concept of quality, if indeed they think about it at all. The protagonist in Robert Pirsig’s book *Zen and the Art of Motorcycle Maintenance* confronts the idea of research quality in academic writing and comes away empty-handed: “...you know what it is, yet you don’t know what it is”, he says (Pirsig, 1975, 178). Further, he says, it is generally accepted that some things are better than others, that is, that some things are of higher quality than others. “But if you can’t say what Quality is, how do you know what it is, or how do you know that it even exists? If no-one knows what it is, then for all practical purposes it doesn’t exist at all. But for all practical purposes it really *does* exist. What else are the [students’] grades based on? Why else would people pay fortunes for some things and throw others in the trash pile?” (Pirsig, 1975, 178). What, then, is this thing called quality? And when we find it, how can we measure it?

A review of the literature suggests that there are at least four different approaches to the identification and assessment of quality in research reporting:

- *Quality as ultimate good.* This is based on the Platonic idea, expressed, for example, in *The Republic*, of quality as “an absolute beauty and an absolute good”. While in the real world everything may fall short of true quality, it is nevertheless something at which we might aim. In our postmodern world, absolutes are greeted with suspicion, and there is little work in the fields of library and information science or education, for example, in recent years, that assumes a reference point of absolute quality or “perfection”.
- *Quality as relative value.* In this approach, research reports are assessed against each other, and found to be relatively superior, average or inferior in relation to the collection of reports as a whole. This type of evaluation may be carried out for a particular purpose, for example, for selecting papers for a research conference, or for inclusion in a book.
- *Quality as a social or cultural construct.* Quality may have different meanings in different contexts, cultures or countries, and for different people. For example, a Danish study of research evaluation (Hansen, 1996) stressed the different aims of research evaluation in different countries (in one country, research evaluation was aimed at controlling research output, and in another at supporting the development of research productivity and quality) and the different processes used for evaluation in the different countries as a result.

- *Quality based on criterion-referenced evaluation.* Quality assessment becomes an assessment of the extent to which a research report meets certain pre-established criteria, those criteria often being related to the purposes of the evaluation. Criteria for assessing the quality of research are outlined in a number of studies. Smith (2001), for example, suggests that “relevance, timeliness, objectivity, and methodological integrity are the hallmarks of high quality research in applied fields” such as library and information science, while Calvert and Zengzhi (2001) used six criteria for evaluating research articles (with articles being rated on a ten-point scale for each). Models like READER and CriSTAL, described above, are based on this kind of approach.

The literature review shed some light on ideas about research quality and the ways in which research quality has been perceived and measured, across a number of fields and within a number of different theoretical frameworks. It provided some insights into the problems associated with making judgements about research quality and the attempts that have been made to address these problems. However, it also raised many questions and suggested some research challenges that have been taken up in my current research programme.

A pilot study was designed as a preliminary attempt to investigate issues associated with determining the quality of published research in the field of library and information science; a detailed report has already been published (Clyde, 2004). The pilot study had two main aims: first to test the extent to which experienced research evaluators agreed in their ranking of research articles on the basis of quality; and, secondly, to investigate approaches to evaluation used by these experienced evaluators. The underlying assumption in the selection of participants was that we might be able to learn from the experience of people who evaluate research on a regular basis.

A qualitative, naturalistic research design was used, with participants being asked to perform a task associated with research evaluation and then to comment on the task. Specifically, they were asked to evaluate a set of five published research articles by ranking them according to their quality (as the participants understood the concept), and then commenting on their reasons for choosing the article they ranked first. They were also invited to reflect on the process of ranking the articles. The five articles had been selected at random from my own comprehensive database of research articles and papers in the field of school librarianship published in English since 1991 (a database created to support my long-term “Research and Researchers in School Librarianship” project, see Clyde, 2002). The evaluators fell into two distinct groups: people with experience of evaluating research in the specific field of school librarianship (and with knowledge of school librarianship as a topic); and people with experience of evaluating research *per se* (and who have no specialist knowledge of school librarianship as a topic). The evaluators included editors of international research journals, professors whose teaching field is research methodology, referees for major research conferences, and assessors for research databases, among others. The data analysis strategies included a visual inspection of the rankings assigned to the articles by the evaluators, along with statistical analyses of the rankings, and content analysis of the detailed comments made by the evaluators.



On the basis of the analysis of the rankings, one of the findings that emerges most clearly is that even people who have a great deal of expertise in the evaluation of research reports, will disagree in their rankings of research reports, to the point where what is ranked first by one expert may be ranked lowest by another. Every one of the five articles was ranked first by at least one evaluator. Further, the experts may claim to be using the same criteria or the same strategies when coming to these very different conclusions. These judgements are not subjective: the experts can justify their rankings and provide evidence for the ways in which they reached their conclusions. If experts evaluating a research report can come to very different conclusions about it, then less experienced evaluators may find the evaluation of research (for example as the basis of evidence-based practice) a confusing process.

Cluster analysis was used to identify groupings or clusters of evaluators according to the way in which they ranked the articles. This statistical analysis suggested that differences (tentatively called “value perceptions”) among evaluators may be even more important than specialist subject knowledge or other factors. Three clusters of evaluators emerged from this analysis; the clusters were confirmed through a content analysis of the comments provided by the evaluators. The clusters were as follows:

- *Cluster 3.* Evaluators in this group placed strong emphasis on empirical or “scientific” research. They particularly valued the provision in a research report of sufficient information to enable an evaluator to see exactly how the research was done and another researcher to repeat the work in order to evaluate it. This was true whether the research was quantitative or qualitative.
- *Cluster 2.* Evaluators in this group were particularly concerned with factors external to the research itself but having an impact on it — for example, ethical issues, the value of the research to the profession or the wider community, and the timeliness of the research. These evaluators liked to see significant questions being addressed, in a timely manner, and using a fresh approach.
- *Cluster 1.* Evaluators in this group emphasized one (or commonly both) of two factors: whether or not the research met the implicit or explicit criteria used by the evaluator to assess quality; and/or whether or not the research was set within a context of the literature and previous research (itself a criterion).

More research would be needed to arrive at a better understanding of these “value perceptions”, but it is clear that they are the things that the evaluators hold to be most important when looking at research reports. We don’t know yet how these “value perceptions” are formed, nor do we know how they are related to other concerns of the evaluator. It is possible that other factors influence people who are evaluating research reports. We know, for example, that knowledge of the subject field plays a part, though perhaps not a crucial one; each of the three clusters included evaluators with specialist knowledge of school librarianship and those without such specialist knowledge.

## CONCLUSION

If research is to inform practice in our field, then practitioners need to be able to evaluate research reports and interpret the results in a way that will be helpful within their own setting. However, there are many problems associated with evaluating research, not least of which is the different understandings about quality that people bring to the evaluation process. A number of strategies and models have been proposed for evaluating the quality of published research (particularly in the form of research articles), and almost all of these strategies or models are backed by some research. However, all have been found to have some limitations or to be unsatisfactory in some way, particularly for evaluating research to support practice in a setting where decisions need to be made quickly. Because research quality means different things to different people and in different settings, developing a single measure of research quality may be difficult, even assuming that just one measure or strategy for all purposes, is desirable. My research (described above), based on the ways in which experts in research evaluation, actually evaluate research reports, suggests that research evaluation is a much more complex process than we usually acknowledge. This is a rather inconvenient finding for those who believe that relatively simple measures like citation counts or impact factor analysis by themselves actually provide meaningful information about the quality of individual research reports.

This research is continuing, with two analyses in progress as this paper was being written. The first is a more detailed analysis of the evaluation strategies used by participants in the pilot study. The second is an analysis of responses provided by participants in a questionnaire survey of active researchers in the field of school librarianship. The latter will provide information about how these researchers see research evaluation and the strategies used in their institutions for research evaluation.

## REFERENCES

- (2004), "ABC sore as Saddle Club left out of reckoning", *The Australian*, 15 July, p.24.
- Audunson, Ragnar (2004), "Is that really so? Some guidelines when evaluating research". Paper presented at the World Library and Information Congress: 70<sup>th</sup> IFLA General Conference and Council, 22-27 August 2004, Buenos Aires, Argentina. Available at <http://www.ifla.org/IV/ifla70/papers/094e-Audunson.pdf> (accessed 9 May 2005).
- Avkiran, N.K. (1997), "Scientific collaboration in finance does not lead to better quality research", *Scientometrics*, 39(2): June, pp.173-184.
- Bahr, A.H. and M. Zemon (2000), "Collaborative authorship in the journal literature: Perspectives for academic librarians who wish to publish", *College and Research Libraries*, 61(5): September, pp.410-419.
- Blake, V.P.L. (1996), "The perceived prestige of professional journals, 1995: A replication of the Kohl-Davis study", *Education for Information*, 14: pp.157-179.
- Bleakley, C. and D. MacAuley (2002), "The quality of research reports in journals", *British Journal of Sports Medicine*, 36(2): April, pp.124-125.

- Booth, A. (2003), *CriSTAL Online*. Available at <http://www.shef.ac.uk/~scharr/eblib/> (accessed 22 January 2004).
- Booth, A. and A. Brice (2003), "Clear-cut? Facilitating health librarians to use information research in practice", *Health Libraries Review*, 20(1): June, pp.45-52.
- Calvert, Philip J. and Shi Zengzhi, (2001), "Quality versus quantity: Contradictions in LIS journal publishing in China", *Library Management*, 22(4-5): pp.205-211.
- CASP (2003), *The Critical Appraisal Skills Programme (CASP) and Evidence-Based Practice*. Available at <http://www.phru.nhs.uk/~casp/casp.htm> (accessed 22 January 2004).
- Clyde, Laurel A. (2002), "Developing the knowledge base of the profession: Research in school librarianship". In D. Singh, A. Abdullah, S. Foneska, and B. de Rozario (Eds), *School Libraries for a Knowledge Society: Proceedings of the 31<sup>st</sup> Annual Conference of the International Association of School Librarianship.... Petaling Jaya, Malaysia, 5-9 August 2002*. Seattle, WA: International Association of School Librarianship, pp.55-75.
- Clyde, Laurel A. (2004), "Evaluating the quality of research publications: A pilot study of school librarianship", *Journal of the American Society for Information Science and Technology*, 55(13): November, 1119-1129.
- Day, A., and J. Peter (1994), "Quality indicators in academic publishing", *Library Review* 43(7): pp.4-72.
- Derbyshire, David (2004), "Researchers decry hormone therapy study that alarmed women", *Sydney Morning Herald*, 29 June, News p.9.
- Eldredge, J. (1997), "Evidence-based librarianship: A commentary for *Hypothesis*", *Hypothesis*, 11(3): 4-7.
- Fazackerley, A. (2003), "Call to name and shame review flouters", *The Times Higher Education Supplement*, 15 August, 2-3.
- Feinstein, Alvan R. and Ralph I. Horwitz (1997), "Problems in the 'evidence' of 'evidence-based medicine'", *The American Journal of Medicine*, 103(6): December, pp.529-535.
- Garfield, E.E. (1979), *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*, New York, NY: John Wiley.
- Gehanno, J. and B. Thirion, (2000), "How to select publications on occupational health: The usefulness of MEDLINE and the impact factor", *Occupational and Environmental Medicine*, 57: pp.706-709.
- Giles, M., F. Mizell and D. Patterson (1989), "Political scientists' journal evaluation revisited", *PS: Political Science and Politics*, 22: pp.613-617.
- Gorman, G.E. and Philip J. Calvert (2000), "Journal quality in the Asian region: Results of a pilot study for the IFLA Round Table of Library and Information Science Journals". Paper presented at the 66<sup>th</sup> IFLA General Council and Conference, Jerusalem, August 2000. Available at <http://www.ifla.org/IV/ifla66/papers/156-125e.htm> (accessed 9 May 2005).
- Hansen, H.F. (1996), "Evaluation of research: The Danish experience in international perspective", *Biblioteksarbejde*, 48: 7-18.

Hart, R.L. (undated), "Collaborative publications by university librarians: An exploratory study", Unpublished manuscript, cited in A.H. Bahr and M. Zemon (2000), "Collaborative authorship in the journal literature: Perspectives for academic librarians who wish to publish", *College and Research Libraries*, 61(5): September, pp.410-419.

Harter, S.P. (1996), "The impact of electronic journals on scholarly communication: A citation analysis", *Public-Access Computer Systems Review*, 7(5). Available at <http://info.lib.uh.edu/pr/v7/n5/hart7n5.html> (accessed 29 January 2003).

Haycock, Ken (1994), "Research in teacher librarianship and the institutionalization of change", in 23<sup>rd</sup> *International Conference, International Association of School Librarianship, Selected Papers, School of Library and Information Science, University of Pittsburgh, Pittsburgh, Pennsylvania, July 1994*, Pittsburgh, PA: International Association of School Librarianship, 94-103.

Henderson, Mark (2003), "MMR doctor accused of misconduct by MP", *The Times*, 17 March, News p.2.

Hernon, P. (1999), "Research in library and information science — Reflections on the journal literature", *The Journal of Academic Librarianship*, 25(4): 263-266.

Humes, Walter and Tom Bryce (2001), "Scholarship, research and the evidential basis of policy development in education", *British Journal of Educational Studies*, 49(3): September, pp.329-352.

Institute of Scientific Information (2002), Journal Selection Criteria. Available at <http://www.isinet.com/isi/journals/jsel.html> (accessed 29 January 2003).

Institute of Scientific Information (2005), ISI Master Journal List. Available at <http://www.isinet.com/cgi-bin/jrnlst/jloptions.cgi?PC=master>

Justice, A.C., *et al.* (1998), "Does masking author identity improve peer review quality?", *Journal of the American Medical Association*, 280: pp.240-242.

Katzer, J., K.H. Cook and W.W. Crouch (1998), *Evaluating Information: A Guide for Users of Social Science Research*, 4<sup>th</sup> edn, Boston, MA: McGraw-Hill.

Kirkey, Sharon (2004), "Study questions anti-depressant drugs for kids", *The Province*, 9 April, p.A2.

Kohl, D.F. and C. Davis, (1985), "Ratings of journals by ARL directors and deans of library and information science schools", *College and Research Libraries*, 46(1): pp.40-47.

Kostoff, R.N. (1998), "The use and misuse of citation analysis in research evaluation", *Scientometrics*, 43(1): September, pp.27-43.

Lee, K.P., M. Schotland, P. Bacchetti, and L.A. Bero (2002), "Association of journal quality indicators with methodological quality of clinical research articles", *The Journal of the American Medical Association*, 287(21): pp.2805-2808.

Lowi, T.J. (1992), "The state of American political science", *American Political Science Review*, 86: pp.210-215.

- MacAuley, D. (1994), "READER: An acronym to aid critical reading by general practitioners", *British Journal of General Practice*, 44: pp.83-85.
- MacAuley, D. and E. McCrum (1999), "Critical appraisal using the READER method: A workshop-based controlled trial", *Family Practice*, 16(1): February, pp.90-93.
- MacAuley, D., E. McCrum and C. Brown (1998), "Randomised controlled trial of the READER method of critical appraisal in general practice", *British Medical Journal*, 316(7138): 11 April, pp.1134-1137.
- MacRoberts, M.H. and B.R. MacRoberts (1989), "Problems of citation analysis: A critical review", *Journal of the American Society for Information Science*, 40(5): pp.342-349.
- McClure, Charles and A. Bishop, (1989), "The status of research in library / information science: Guarded optimism", *College and Research Libraries*, 50: 127-144.
- McQuay, Henry (2003), "Guessing games with large needles", *The Times Higher Education Supplement*, 15 August, p.25.
- Moed, H.F. (2002), "The impact-factors debate: The ISI's uses and limits", *Nature*, 415: 14 February, pp.731-732.
- Nicolaisen, J. (2002), "The J-shaped distribution of citedness", *Journal of Documentation*, 58(4): pp.383-395.
- Nkerekwem, E.E. (1997), "Accrediting knowledge: The ranking of library and information science journals", *Library Review*, 46(2): pp.99-104.
- Opthof, T. (1997), "Sense and nonsense about the impact factor", *Cardiovascular Research*, 33, pp.1-7.
- Pálsdóttir, Guðrún, *et al.* (1997), "Published research about library and information science in or related to Iceland", *Nordic Yearbook of Library, Information and Documentation Research*, Oslo: Novus forlag, 85-108.
- Peters, D.P. and S.J. Ceci (1982), "Peer review practices of psychological journals: the fate of published articles, submitted again", *Behavioral and Brain Sciences*, 5: pp.187-255.
- Pirsig, Robert M. (1975), *Zen and the Art of Motorcycle Maintenance*, London: Bantam.
- Plato (360 BC), *The Republic* (translated by B. Jowett). Available at <http://classics.mit.edu/Plato/republic.7.vi.html> (accessed 9 September 2002).
- Reuters (2005), "MIT students pull prank on conference", CNN.com, 14 April. Available at <http://edition.cnn.com/2005/TECH/science/04/14/mit.prank.reut/index.html> (accessed 15 April 2005).
- Ross, S. (2002), "Becoming a good consumer of evidence", *Principal Leadership*, 3(2): pp.10ff.
- Seglen, P.O. (1997), "Why the impact factor of journals should not be used for research evaluation", *British Medical Journal*, 314: pp.498-502.
- Smith, A. (2001), "Never mind the width, Feel the quality: Improving VET research in Australia". In *Research to Reality: Putting VET Research to Work. Proceedings of the*

*Australian Vocational Education and Training Research Association Conference, Adelaide, Australia, March 2001.* Available at <http://www.avetra.org.au/PAPERS%202001/a%20SMITH.pdf> (accessed 9 September 2002).

Sokal, Alan (1996), "Transgressing the boundaries: Toward a transformative hermeneutics of quantum gravity", *Social Text*, 46-47: Spring/Summer, 217-252.

Thomas, O., and P. Willet (2000), "Webometric analysis of departments of librarianship and information science", *Journal of Information Science*, 26(6): pp.421-428.

Todd, Ross J. (2003), "Irrefutable evidence", *School Library Journal*, 49(4): April, p.52.

Tooley, J. and D. Darby (1998), *Educational Research: A Critique*, London: HMSO.

Vaughan, L. and K. Hysen (2002), "Relationship between links to journal web sites and impact factors", *ASLIB Proceedings*, 54(6): pp.356-361.

Viadero, D. (2002), "Researching the researchers", *Education Week*, 21(23): pp.26-29.

Warner, J. (2000), "A critical review of the application of citation studies to the Research Assessment Exercises", *Journal of Information Science*, 26(6): pp.453-459.

Wood, Michael and Martyn Roberts (2004), "The reliability of peer reviews of papers on information systems", *Journal of Information Science*, 30(1): pp.2-11.

---

## **Biographical Information**

Dr L. Anne Clyde is Professor in the Library and Information Science Department at the University of Iceland, Reykjavík. An Australian citizen, she worked in universities in her home country and in Canada before moving to her present position. Her books include *School Libraries and the Electronic Community: The Internet Connection* (1997), *Managing InfoTech in School Library Media Centers* (1999), and *Weblogs and Libraries* (2004). She has contributed articles to a number of international research journals, including *Library and Information Science Research*, *Journal of the American Society for Information Science and Technology*, *The Electronic Library*, and *Library Management*. She is currently Chair of the IFLA Section of School Libraries and Resource Centres.